Market reaction to the positiveness of annual report narratives

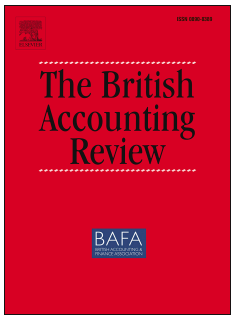
Yekini, S., Wisniewski, T. P. and Millo, Y.

Original citation & hyperlink:
Yekini, S. , Wisniewski, T. P. and Millo, Y. (2015) Market reaction to the positiveness of annual report narratives. *The British Accounting Review*, volume 48 (4): 415-430
http://dx.doi.org/10.1016/j.bar.2015.12.001

DOI 10.1016/j.bar.2015.12.001
ISSN 0890-8389

Publisher: Elsevier

Accepted Manuscript



Market Reaction to the Positiveness of Annual Report Narratives

Liafisu Sina Yekini, Tomasz Piotr Wisniewski, Yuval Millo

PII: S0890-8389(15)30004-4

DOI: 10.1016/j.bar.2015.12.001

Reference: YBARE 712

To appear in: *The British Accounting Review*

Received Date: 3 January 2015

Revised Date: 1 December 2015

Accepted Date: 2 December 2015

Please cite this article as: Yekini, L.S., Wisniewski, T.P., Millo, Y., Market Reaction to the Positiveness of Annual Report Narratives, *The British Accounting Review* (2016), doi: 10.1016/j.bar.2015.12.001.

Market Reaction to the Positiveness of Annual Report Narratives

Liafisu Sina Yekini *
Coventry University
Coventry Business School
William Morris Building
Gosford Street
Coventry, CV1 5DL, UK
Tel: +44 024 7688 8440
E-mail: sina.yekini@coventry.ac.uk

Tomasz Piotr Wisniewski
University of Leicester
School of Management
Ken Edwards Building
University Road
Leicester LE1 7RH, UK
Tel: +44 116 252 3958
E-mail: t.wisniewski@le.ac.uk

Yuval Millo
University of Leicester
School of Management
Ken Edwards Building
University Road
Leicester LE1 7RH, UK
Tel: +44 116 229 7385
E-mail: ym95@le.ac.uk

* Corresponding author

Market Reaction to the Positiveness of Annual Report Narratives

Abstract

This paper focuses on narratives published by UK companies, defined here as the content of annual reports excluding financial statements and notes to accounts. We endeavour to gauge the tone of these narratives by recording the frequency of positive words appearing in the text. We show that the extent of positiveness is related to market reaction around the disclosure date. This conclusion is maintained even after controlling for the financial figures that are reported simultaneously and company-specific characteristics. Consequently, narratives should not be perceived as mere impression management tools, but also as conduits for disseminating price-sensitive information.

JEL codes: M41; G12; G14

Keywords: Annual report narrative, content analysis, stock market reaction

1. Introduction

Recent academic studies have paid increasing attention to the market impact of the tone inherent in corporate narratives (Frankel, Mayew, & Sun, 2010; Davis & Tama-Sweet, 2012; Huang, Teoh, & Zhang, 2014). Previous literature examined, amongst others, corporate press releases (Ahern & Sosyura, 2014), quarterly earnings announcements and conference calls (Francis, Schipper, & Vincent, 2002; Demers & Vega, 2010; Price, Doran, Peterson, & Bliss, 2012; Davis, Piger, & Sedor, 2012). Notwithstanding the advances made in this area of research, relatively little attention has been paid to the narratives of annual reports as a whole. The extant content analysis of annual reports has been predominantly restricted to relatively short sub-sections, such as company presidents' letters to shareholders (McConnell, Haslem, & Gibson, 1986; Abrahamson & Amir, 1996; Swales, 1988; Smith & Taffler, 2000). A study by Loughran and McDonald (2011), who linked the tone of full 10-K forms to filling period returns, constitutes a notable exception. This gap in the literature is significant, as a growing body of work suggests that the language used in the qualitative part of annual reports plays a role in framing the quantitative section of the documents (Hoskin, Hughes, & Ricks, 1986; Anilowski, Feng, & Skinner, 2007). Although both qualitative and quantitative information are often released concurrently, recent research has employed statistical techniques to separate the influence of these two different modes of representation (Francis *et al*., 2002; Demers & Vega, 2010). We use the same methodological approach in this paper. A better understanding of the potential impact of the narratives in annual reports will not only help us to understand this aspect of corporate communications, but also shed more light on the overall impact of these documents on markets.

While we acknowledge the contribution made by Loughran and McDonald (2011), our research expands their work by examining annual reports published by UK rather than US companies. Due to the different regulatory frameworks, one should not expect that the results

reported in Loughran and McDonald (2011) will be applicable also in the UK context. Notably, the accuracy and complexity of communications may be dissimilar under principles- and rules-based standards (Nelson, 2003). Schleicher and Walker (2010) also note that disclosure-related litigation is significantly less common in the United Kingdom than in the United States, while Frost and Pownall (1994) and Davies (2009: 311) remark on the SEC's greater stringency in monitoring and enforcing annual and interim reporting rules. During the 2005-2014 period, there were 1,300 SEC enforcement actions related to issuer reporting and disclosure (SEC, 2015). In the UK, on the other hand, the laws against fraudulent misstatements and misleading reporting do not appear to be sufficiently deterrent (Ferran, 2009: 343) and the definition of fraud is narrower than in the US (Davies, 2007: 44). If legal considerations were paramount to reputational issues, one would expect that UK managers would be more inclined to manipulate the tone of annual report narratives.

It must be acknowledged that a number of previous studies employed content analysis based on UK data. Smith and Taffler (1995) examined whether their research participants were able to discriminate between failed and non-failed companies by reading chairmen's statements. They also conducted additional investigations inquiring into whether keywords and themes contained in chairmen's statements, as well as readability and understandability of these texts, predict bankruptcy (Smith & Taffler 1992, 2000). Schleicher and Walker (2010) employed a manual analysis method to scrutinize the tone of the outlook section, which is a part of the chairman's statement. They found that firms with impending declines in sales growth and profit margin bias the tone upwards and that the tone depends on managerial incentive variables. Clatworthy and Jones (2003) look at how chairmen's narratives relate to profit before taxation and find that management tends to report in a way that best serves its own interests, crediting itself for good news and attributing bad news to the external environment. Finally, Brennan,

Guillamon-Saorin, and Pierce (2009) focus on creating a new multi-dimensional impression management measure.

While these papers make a substantial contribution to our understanding of qualitative managerial communications, they differ substantially from the current paper. Ours is a study that, unlike its UK predecessors, examines market reaction (as measured by returns) to the disclosure of narratives. In doing so, it also focuses on longer documents. This is particularly helpful, as identification of linguistic style is more reliable for lengthy texts (Grimmer & Steward, 2013: 272). Such observation has also been made in the field of authorship attribution analysis, where longer textual samples are considered to have greater discriminatory power in terms of distinguishing authorship characteristics (Baillie, 1974; Ledger & Merriam, 1994). Although all these differences between our paper and prior studies are non-trivial, there is one striking similarity worth mentioning. Namely, we found that narratives disclosed by companies are indeed an important medium of communication and that they should not be overlooked.

The overarching aim of this paper is to measure the degree of positiveness woven into annual report narratives of UK companies between January 2006 and June 2013. The 'narrative', to which we also refer as the 'qualitative part of an annual report', is defined here as the content of the annual report including independent auditors' report, but excluding financial statements and the notes to the accounts. Following the accepted convention, any numbers appearing in the narrative are not removed from the text. Furthermore, from a definitional perspective, in this study we use the word 'tone' when referring to positiveness or net positiveness of narratives. We focus on gauging the impact of positiveness because managers may either be incentivized to send an upbeat message to investors and other users of accounts or, alternatively, they may simply be expressing their understanding of corporate performance. An ongoing debate in accounting literature revolves around the question of whether qualitative disclosure is a vehicle for incremental information content or an element of managerial attempts

to carry out impression management (Merkl-Davies, Brennan, & McLeay, 2011). We believe that our study can contribute to this debate by demonstrating that the tone of narratives indeed conveys information, which investors find to be material in nature. Consequently, it is imperative that market participants familiarize themselves with these narratives as soon as they become available. We also want to note at this stage that measuring market reaction is important as it allows us to make inferences about investor behaviour.

Our findings show that abnormal returns in short windows surrounding financial report disclosure dates correlate significantly with the positiveness encapsulated in annual reports. However, deriving inferences from a simple correlation coefficient can be misleading, due to the fact that financial figures are disclosed concurrently with narratives. The disentanglement of the individual influences exerted by the narratives and the quantitative data has been attempted here in a multivariate regression framework. The results obtained from this analysis indicate that positiveness remains an important factor explaining market reaction, even after financial performance and company characteristics have been controlled for.

The remainder of the article is organized as follows. The next section reviews the literature in the field and provides justification for our study. This is followed by a section on theoretical considerations and a further one describing the methodological framework. Section 5 elaborates on our data sources and the characteristics of the sample, while Section 6 presents the results of our empirical analysis. Section 7 offers additional considerations and the paper ends with some conclusions and a list of implications.

2. Research Context

Our investigation employs content analysis in the context of corporate annual report narratives. One way to approach this task is to rely on human judgment to evaluate these texts.

However exploring this avenue would be extremely time-consuming given that our sample comprises 1,410 lengthy documents. There are a number of researchers who took this route, including Bhattacharya, Galpin, Ray, and Yu (2009) who read and evaluated over 171 thousand news items about Internet IPOs, and Smith and Taffler (1995) who engaged 146 students to process chairmen's statements. A second method is to employ statistical evaluation of textual data – an approach that was introduced in the accounting context by Frazier, Ingram, and Tennyson (1984) who used a software prototype called *Words*. Since then, there has been considerable progress in the software applications available and in quantitative linguistics, however most of the algorithms rely on computing the frequencies of words falling into a given category.

In our research we rely on the thesaurus of positive words developed by Henry (2008: 387) that was created specifically for accounting and finance applications. Henry's wordlist has already gained some popularity among researchers. For instance, it has been used to measure the tone of R&D disclosures (Merkley, 2014), transcripts of earnings conference calls (Doran, Peterson & Price, 2012; Price *et al.*, 2012; Davis, Ge, Matsumoto & Zhang, 2015), discretionary disclosures prior to restatements (Gordon, Henry, Peytcheva & Sun, 2013) and investment proposals submitted to business angles (Parhankangas & Ehrlich, 2014). Rogers, Van Buskirk and Zechman (2011) employ Henry's library of words to document that the use of overly optimistic language in earnings announcements increases litigation *risk.*

Importantly, extant literature suggests that the tone of qualitative corporate reports may have non-negligible market impact. Kothari and Short's (2003) findings indicate that positive disclosure favourably affects cost of capital and price volatility, while Li (2010) shows that the tone of forward-looking statements has predictive power for the company's future performance. Furthermore, Davis *et al.* (2012), Demers and Vega (2010) and Huang *et al.* (2014) documented

that an optimistic disposition in corporate earnings press releases is associated with abnormal market returns following the announcements.

Following this trajectory in the literature, we identified two important points that affected our research design, namely the focus on the qualitative part of annual reports and the attention to positive tone within the text. In contrast with much of the existing literature - where a variety of corporate narrative outputs, such as earnings announcements, press releases, as well as texts from financial mass media (Tetlock, 2007; Tetlock, Saar-Tsechansky, & Macskassy, 2008) are examined - we decided to focus our examination on the qualitative part of companies' annual reports. In the UK, the Companies Act 2006 and the amendments to this Act introduced in 2013 require large and medium listed companies to incorporate certain sections in their annual reports. These include the strategic report/business review section (covering business description, issues related to performance, principal risks, position, trends and factors, and key performance indicators), the corporate social responsibility statement (describing environmental, employee and community issues), the directors' report, the directors' remuneration report, and the statement of directors' responsibilities. The UK Corporate Governance Code provides guidance on the directors' remuneration and directors' reports, while the Listing Rules require companies to either comply with the UK Corporate Governance Code or explain why they have failed to do so. Disclosure and Transparency Rules outline a framework for providing corporate governance statements. Compliance with regulations is monitored by the Financial Reporting Council.[1] While managers are still afforded significant discretion as to how to frame corporate performance and how much optimism to inject into the narrative, texts produced by different companies comprise a relatively structurally homogenous sample, which contributes to more effective analysis of the examined factor in the text.

[1] This obligation arises from the regulatory need to enforce the Companies Act 2006 and the amendments to this act introduced in 2013. This regulatory body also issued guidance on strategic and directors' reports.

Our focus on positiveness follows from the contextual nature of qualitative texts. Although we acknowledge the issue of inherent context-dependence of qualitative corporate disclosures, our measurement choice incorporates the assumption that qualitative parts in annual reports are aimed at communicating an overall positive tone and that they do so by repeatedly signalling to the reader the positiveness of the firm's activities, regardless of the immediate context in which these are presented. In the case of such texts, which are contextually positive, a measure of the degree of upbeat tone, above and beyond what investors already accept, may affect investors' decision-making. In fact, the reaction of market participants is likely to be more pronounced if they perceive the narrative part of annual reports to be a conduit for new material information rather than a mere impression management tool.

3. The Influence of Narratives

To gain a better understanding of the association between the positiveness of narratives and market response, we deploy a theoretical framework that addresses the potential underpinning mechanisms of this phenomenon. Research focusing on the economic utility that investors receive from qualitative disclosures finds that tone conveys decision-relevant information. For example, Li (2010) finds an association between tone and future earnings, while Davis *et al.* (2012) show that managers use language to signal expectations about the firm's performance. We wish to theorise how this information is incorporated and utilised in the investor's decision-making, which we regard in this context to be primarily a cognitive process. That is, although the environment in which decisions are taken may sometimes be comprised of groups of investors, the activity of reading the narrative is essentially an individual cognitive process. Following this, we mobilize theories from cognitive psychology that posit that the tone incorporated in managerial qualitative disclosures, such as those contained in annual reports, influences attitude change (Ajzen & Fishbein, 2000; Crano & Prislin, 2006; Perloff, 2010).

The influence potential of texts is regarded primarily in psychology as informational influence, which is seen as cognitive responses to persuasive messages (Petty, Cacioppo, & Schumann, 1983). According to the Elaboration Likelihood Model (ELM) put forward by Petty and Cacioppo (1986), when actors are exposed to potentially persuasive communication and when they are attentive to this message, they engage in a cognitive structure change. This process encourages the creation of new cognitions that may then be adopted and stored in memory and, as a result, different responses become more salient than previously (Sussman & Siegal 2003; Petty & Cacioppo 2011). Put differently, influence is conceptualized as a cumulative series of signals in the text that, when read by the actor, gradually brings about cognitive structure change and a change in attitude.

Further findings from experimental psychology and marketing are also relevant. Wilson and Miller (1968) document that repeated arguments become more persuasive, while Weiss (1969) finds that repeated exposure to a message is related to a higher degree of opinion formation. Commenting on the existing literature, Malaviya, Meyers-Levy, and Sternthal (1999) note that increasing the number of exposures to commercial advertising affects its persuasiveness. The literature acknowledges that the association between repetition and persuasion may be moderated by a number of factors, such as credibility of source or argument strength.

In the light of ELM, we conceptualize investors reading the narratives in the firms' reports with the aim of deciding whether or not they should invest in, or alternatively, remove their investment from the firms in question as a case of engagement in the process of cognitive structure change. Thus, the relation between repetition and attitude change emphasised in the literature also motivates us to consider frequency as an important factor in examining the market impact of annual report narratives. That is, the more often positive tone expressions are mentioned in the text, the more likely it is that they will be influential. Importantly, we do not

theorize directly about the mechanisms that attribute market reaction to the narratives, but our reading of the literature on informational influence directs our research design. In particular, it provides justification for choosing a measurement tool that relies on comparing frequencies of positive tone articulations.

We operationalized our inquiry by programming Henry's (2008) positiveness thesaurus into a content analysis software application called Diction.[2] By doing so, we were able to measure the frequency with which positive words are found in each of the annual report narratives. These frequencies were subsequently linked to abnormal returns around disclosure dates. We have examined two different event windows and two different statistical models against which abnormal returns are defined. Our null hypothesis is that market reaction is *ceteris paribus* unrelated to the tone inherent in the narratives. A range of control variables has been collected and incorporated into the regressions in order to more cleanly isolate the influence of tone.

4. Methodology

Since one of our main objectives is to measure market reaction around a specific event, that is the publication of an annual report, we employ event-study analysis which is suitable for the task at hand (Brown and Warner, 1980; 1985). Its aim is to measure abnormal returns (*AR*s) that are directly attributable to certain occurrences. An *AR* is defined as a deviation of the observed return from the return that would have materialized in absence of the event. Of course,

[2] Diction has become a very popular software package and has found many applications in political science, communication studies, linguistics, business studies, and sociology. Creators of the software track all publications that used it and list them on the following website http://www.dictionsoftware.com/published-studies/ . At the time of writing this paper, Diction was used by the authors of 151 refereed journal articles, 15 books and monographs, 58 book chapters, 68 conference presentations, as well as 49 working papers and proceedings. Notable examples of applications of the Diction software in the field of accounting include Sydserff and Weetman (2002), Yuthas, Rogers and Dillard (2002), Demers and Vega (2010), Rogers *et al.* (2011), Craig and Brennan (2012), and Davis *et al.* (2012).

there is no way of knowing with certainty what would have happened had the event not occurred. For this reason, instead of relying on an unknown hypothetical construct, *AR*s typically gauge returns in excess of some pre-determined statistical benchmark.

As Campbell, Lo, and MacKinley (1997) note, there are two commonly adopted benchmarks in the context of event-study analysis. The first assumes that, under the null hypothesis of no market reaction, the returns for a given event *i* are constant over time and equal to μ_i. This implies that, during the period surrounding the event, the mean-adjusted *AR*s can be defined as follows:

$$AR_{i,t}^{MA} = R_{i,t} - \hat{\mu}_i \qquad [1]$$

where $R_{i,t}$ is the return on the relevant company for event *i* observed on day *t* and where $\hat{\mu}_i$ has been estimated as an average return on that company during a period preceding the event window. The second methodological approach accounts for the systematic risk of a security and overall stock market fluctuation. It estimates a single factor model, where stock returns are regressed against a stock market index R_M, as follows: $R_{i,t} = \alpha_i + \beta_i R_{M,t} + \varepsilon_{i,t}$. The estimation is based on data recorded immediately prior to the event window and FTSE350 approximates here the market portfolio. Collecting the parameter estimates, we are able to compute market-model-adjusted *AR*s in the temporal proximity of the event:

$$AR_{i,t}^{MM} = R_{i,t} - (\hat{\alpha}_i + \hat{\beta}_i R_{M,t}) \qquad [2]$$

The timeframe we select to estimate our event-specific parameters (μ_i, α_i, and β_i) begins 201 trading days before the event and ends 2 days before the event. In other words, we use a 200-day estimation window (-201,-2), relative to the first annual report dissemination date (*Day* 0). Had this window been any longer, one would run the risk of the previous year's report disclosure being incorporated within it. Consequently, its usefulness as an event-neutral benchmark would have been invalidated. On the other hand, shortening the estimation span would lead to less precise statistical inferences. With regard to examination of the impact of the

report release itself, we choose to focus on two short periods, namely (-1,1) and (-1,5). Since the windows are relatively narrow, the probability that major confounding events will occur during these ephemeral timeframes is minute, thereby reducing the likelihood of contaminated results. It is worth noting that prior studies focusing on earnings announcements used event windows of comparable length (see for instance Francis *et al.* (2002: 519), Scharnd & Walther (2000: 169)).

In the next step of our analysis, we cumulate the abnormal returns over time for each event within the relevant period to arrive at the cumulative abnormal return:

$$CAR\_X_i(t_1, t_2) = \sum_{j=t_1}^{t_2} AR_{i,j}^{X} \qquad [3]$$

where *X* can take the value of either *MA* or *MM* depending on whether the *AR*s are mean- or market-model-adjusted. The parameter t_1 denotes the beginning of the event window, or *Day -1* in our case, while t_2 can take a value of either +1 or +5 depending on the specification. *CAR_X* can be simply interpreted as the totality of market reaction associated with publication of a particular annual report. In other words, CARs capture the response of investors to the information contained in both the annual financial statements and the narratives.

To disentangle the impact of qualitative information from that of quantitative data and company characteristics, we perform regression analysis. More specifically, our estimate of market response is linked to the positiveness of the descriptive part of the report and other control variables. More formally, we try to fit the following regression to our data:

$$CAR\_X_i(t_1, t_2) = \alpha + \beta_1 Positiveness_i + \beta_2 Size_i + \beta_3 Book\_to\_Market_i + \\ + \beta_4 Earnings\_Surprise_i + \beta_5 \Delta\%Sales_i + \beta_6 \Delta Leverage_i + \varepsilon_i \qquad [4]$$

Detailed definitions of the explanatory variables are provided in Table 1. In our empirical inquiry we try different values of *X* and t_2. Furthermore, in some of the specifications we restrict some of the β coefficients to be equal to zero.

[Table 1 about here]

The key regressor in equation [4] is *Positiveness*, which is defined as the fraction of positive words in the text of an annual report narrative. A large number of studies have utilized variables constructed by dividing the number of words falling into a given tone category by the total number of words in the documents. Examples of papers that operationalized measures of positiveness/optimism constructed in that manner include Feldman, Govindaraj, Livnat, and Segal (2008), Kothari, Li, and Short (2009), Henry (2006), Cicon, Clarke, Ferris, and Jayaraman (2014), Wisniewski and Moro (2014), Ferguson, Philip, Lam, & Guo (2015). Cho, Roertd, and Patten (2010) use the optimism score as a dependent variable to model the biases in corporate environmental disclosures. At this stage it must be mentioned that a number of studies also employ a tone variable, which is typically defined as an unadjusted or scaled difference between positive and negative words (see for instance Henry, 2008; Henry & Leone, 2009; Frankel *et al.*, 2010). We construct our own version of the tone variable and report the results based on it in the Further Considerations section.

The text considered in our study runs from the beginning of the document up to and including the independent auditors' report. Inclusion of auditors' reports is justified on the basis that they include new important information published concurrently with the rest of the annual report and, just like the remainder of the narratives, express opinions that are not subject to a rigorous audit. Financial statements and notes to the accounts are omitted from the calculation of our tone measure, as the contents of these sections has to comply with regulatory requirements and is carefully audited, leaving little scope for linguistic manoeuvring. In our study, we use Henry's (2008: 387) thesaurus of 105 positive words (see Appendix I) that has been developed with the purpose of analyzing texts residing in the domain of accounting and financial reporting. Following Henry (2008) and Henry and Leone (2009), we calculate the frequencies of positive words based on a user-defined dictionary in the computer-aided text analysis program called Diction.

As this wordlist is context-specific and uses specialized language, the problem created by polysemy (words with several different meanings) is partially mitigated.[3] Henry and Leone (2009) document empirically that, when analyzing earnings press releases, this particular word corpus is more powerful than the more general alternatives. This is because it does not include words that are irrelevant in the context of financial disclosure and does not misclassify domain-specific terms. Although the thesaurus has been developed by a US researcher, we believe that it is applicable to the UK market, as it does not include words related to culture or regulations. Furthermore, some British companies may be cross-listed in the US or be willing to attract American capital, which would induce a large degree of language compatibility on both sides of the Atlantic. If the market reacts favourably to positive words printed in reports, one would expect the β_1 coefficient in regression [4] to be positive and statistically significant. Further details on the remaining variables appearing in equation [4] are given in the section that follows.

5. Data

The companies included in our sample are constituents of the FTSE350 stock market index. From the complete list, we have eliminated 72 companies whose operations fell within the financial services domain.[4] Firms that were merged during our sample period were also excluded, as were those with an insufficient number of annual reports or information on financial performance. Our final sample consists of 209 companies listed on the London Stock Exchange. For this group of companies, we manually downloaded available annual reports

[3] The mitigation of the problem of polysemy can be nicely illustrated with the word "beat". In everyday language this word would be associated with violence, while in the context of financial reporting which often refers to beating forecasts or expectations this word may have positive connotations.

[4] Banks and other financial institutions have a different mode of operation compared to other businesses and the financial reporting of these entities is specialized in nature. They have to comply, for instance, with IAS 30 or the Basel Accord and are regulated by the Financial Conduct Authority and Prudential Regulation Authority at the Bank of England. The institutions are subjected to stress testing, have to comply with minimum capital requirements and focus a lot of attention on liquidity and risk management. As a result, the structure of their annual report narratives differs significantly from that of a typical listed company. Some of the other content analysis studies performed for the UK market also concentrate exclusively on non-financial firms (see Clatworthy and Jones, 2003; Schleicher and Walker, 2010).

published between January 2006 and June 2013 from corporate web pages, Morningstar and Bloomberg. At the end of this process we had 1,410 observations, on which more detail is provided in Appendix II to this paper. The publication date for a given report is assumed to be the date on which the report appeared either on Bloomberg or Morningstar, whichever occurred earlier. It should be noted that UK annual reports are almost invariably disseminated as pdf files, which necessitates conversion to the plain text format required for content analysis. The converted files were checked manually to ensure consistency. Finally, we obtained data on company stock prices, market capitalization, book-to-market ratios, and financial indicators from Datastream.

[Table 1 about here]

In addition to cumulative abnormal returns and the *Positiveness* measure, which have already been described in some detail above, this study employs a range of other variables which act as controls (see Table 1). Firstly, the extant literature documents that small companies tend to generate higher returns (Banz, 1981; Fama and French, 1992). Secondly, the seminal work of Rosenberg, Reid, and Lanstein (1985) has discovered a robust relationship between companies' book-to-market ratios and rewards earned by investors. We consequently incorporate the natural logarithm of market capitalization and book-to-market ratios of companies in our set of regressors, which aligns with the argument of Fama and French (1993), who argue that these two can be considered the most important risk factors for stocks. The beta of security is not taken to be an explanatory factor, as the *CAR_MM* dependent variable has already been purged of the influences of the general market.

Furthermore, we try to account for financial figures which are released concurrently with the narrative. Our earnings surprise measure is based on an increase of earnings over a simple random walk forecast. This definition is dictated primarily by data availability and is similar to that used in Wisniewski (2004) and Sponholtz (2008). It should be noted that Hughes and Ricks

(1987) report that an earnings surprise based on a simplistic seasonal random walk benchmark outperforms that derived from analyst forecasts, in that it is more closely linked to abnormal returns around the dissemination date. The numerator of *Earnings_Surprise* is divided by stock price, which coheres with the approach used in Easton and Zmijewski (1989), DeFond and Park (2001), Bartov, Givoly, and Hayn (2002) and Brown and Caylor (2005). We also control for an increase in financial leverage, which according to Bhandari (1988) is *ceteris paribus* positively related to stock returns. Finally, we include the percentage change in sales in our set of regressors. This inclusion is motivated by the findings of Jordan, Waldron, and Clark (2007) who show that sales predict stock prices and Barbee, Mukherji, and Raines (1996) who found that sales yield is one of the strongest determinants of returns.

[Table 2 about here]

Table 2 presents summary statistics for the variables used in our study. The magnitude of *CAR*s is, on average, close to zero. This is not entirely unexpected, as some disclosures will be perceived as good news and others as bad news, cancelling each other out in the averaging process. The mean of our linguistic variable indicates that one in every 209 words appears in our positive tone thesaurus. Furthermore, companies were confronted with falling earnings per share, which can be linked to the occurrence of deep recession during our sample period. The severe impact of the credit crunch is also mirrored in falling financial leverage, as enterprises struggled to access credit due to the banking sector's distress. Even in these difficult circumstances, our sample companies managed to increase their sales volumes, perhaps at the expense of falling profit margins.

[Table 3 about here]

The correlation matrix between variables is reported in Table 3. Most importantly, the Pearson correlation coefficients between positiveness of text and magnitude of market reaction

are positive and statistically significant. This preliminary result attests to the fact that the manner in which annual report narratives are written is not immaterial to stock market participants and that it could possibly convey valuable information. Secondly, correlations between explanatory variables are relatively low, indicating that multicollinearity is not likely to be a problem in the empirical models that follow. In cases where association between the regressors is strong, standard errors of the regression parameter estimates are inflated. Chatterjee and Price (1991) argue that Variance Inflation Factors (VIFs) in excess of 10 are symptomatic of estimation problems. We find that the highest VIF in the regressions reported in this paper is 1.27, dispelling any apprehensions about this potential issue.

6. Empirical Results

In what follows, we analyze different variations of the regression specified in equation [4]. Table 4 reports the results where the dependent variable is defined as the cumulative abnormal return calculated using the mean-adjusted model, while Table 5 focuses on market-model-adjusted CARs. Each of the tables consists of two panels, as two different lengths of the event window are examined. For each panel, three regressions are presented – one with no control variables, one which takes into account company characteristics and one which also incorporates financial performance measures.

[Table 4 about here]

[Table 5 about here]

The most notable finding arising from these tables is that the *Positiveness* measure carries a positive coefficient and is statistically significant in all regression specifications. This has several important ramifications. Firstly, the narrative of annual reports is to some extent flexible and could potentially be manipulated by management. Shin (1994) considers a

theoretical model in which firms that operate in an informationally asymmetric environment could manage the disclosure of facts by suppressing negative news. By the same token, Hildebrandt and Snyder (1981), Rutherford (2005) and Henry (2008) allude to the possible existence of the "Pollyanna effect" (positivity bias) in parts of annual reports. It is conceivable that management will use a report as a marketing tool and suffer from overconfidence when writing their own reviews. The unjustified overuse of optimistic language could undermine the usefulness of information extracted from linguistic features. However, our results indicate that, despite all these real-life complications, the tone of annual reports can still be viewed as price-sensitive in nature.

It is helpful to consider the implications of our findings from the investors' and market's perspectives. According to the Efficient Market Hypothesis proposed by Fama (1970), stock prices already reflect all available information and change only in response to disclosure of previously unknown facts. Consequently, if markets were efficient and the positiveness expressed in narratives was perceived to merely capture different degrees of impression management, rational investors would dismiss it as being uninformative. If, on the other hand, the tone conveyed new important information, discerning market participants would revise their assessment of the company's fundamental value. The discrepancy between fundamental value and market price will induce them to trade. In cases in which the positiveness level turned out to be above expectations, the revised fundamental value would surpass the pre-disclosure market price, leading to a simultaneous lack of supply and excess demand for the company's stocks. The price will need to rise immediately until demand and supply are equalized. In a scenario where the positiveness level was below expectations, the excess of sellers relative to buyers would cause a stock price decline. Several caveats need to be mentioned at this stage. Firstly, for the abovementioned mechanism to work, market participants need to have a general understanding of managerial behaviour and motives, in order to rationally formulate their

expectations of narrative positiveness levels. Secondly, one may argue that the reaction observed can be attributed to noise traders who trade on irrelevant information. However, if this was the case, one would expect smart money to swiftly correct the mispricing that arises as a result. Since the effect of positiveness is not eliminated in our longer event windows, we are inclined to conclude that it is indeed material in nature.

It should also be mentioned that our findings are of interest from the point of view of the UK's regulatory framework. As we have alluded to earlier, UK regulations against potentially misleading disclosure do not appear to be adequately deterrent from the point of view of management, partially due to issuer-only liability (Ferran, 2009). Furthermore, in the US, SEC (1998) provides more detailed guidance on word usage and writing style. Despite this, our finding that the tone of disclosure determines market response is mirrored by those obtained for US earnings press releases (Henry, 2008; Demers & Vega, 2010; Davis *et al.*, 2012). This observation leads us to believe that reluctance to put excessive spin on facts may not be rooted solely in regulatory boundaries, but also in fear of potentially costly reputational loss.

The coefficients on the *Size* variable are always negative and statistically significant for longer event windows, which is consistent with the effect propounded by Banz (1981). There are several reasons why investors may demand higher compensation when committing to low capitalization stocks. Firstly, less information is available on these firms (Atiase, 1985; Freeman, 1987) and analysts are reluctant to follow them (Arbel, Carvell, & Strebel, 1983; Gilbert, Tourani-Rad, & Wisniewski, 2006). Secondly, size is an important determinant of the likelihood of bankruptcy (Shumway, 2001), with large multinationals being more diversified and less risky. Finally, small caps are associated with higher transaction costs (Lesmond, Ogden, & Trzcinka, 1999) and are more strongly affected by the illiquidity problem (Amihud, 2002). Book-to-market ratios, on the other hand, have almost no explanatory power in our regressions, possibly due to the short span of our event windows.

Our findings for *Earnings_Surprise* are in line with those of Lev (1989) who argued that the relationship between accounting earnings and stock market returns is weak and unstable over time. Most likely, financial figures from annual reports do not engender a strong market response because they are merely an aggregation of the interim results that firms have published earlier. This explanation would be in line with Ball and Brown (1968), who observe that about 85 to 90 percent of the information contained in annual report income statements has been captured by reports released beforehand.

One caveat that needs to be mentioned here is that although up to 2007 preliminary statements of annual results announcements were mandatory, they became voluntary due to changes in Listing Rules. We have tested whether this regulatory change had implications for the stability of coefficients on the variables derived from financial statements data (*Earnings_Surprise*, *Δ%Sales*, and *ΔLeverage*). Our tests [not reported] revealed that the null hypothesis of coefficient constancy could not be rejected.[5] Even though disclosure of preliminary statements may have, at least in the latter part of our sample, an element of voluntariness, Disclosure and Transparency Rules (DTR 4.2.2) required companies to disclose half-yearly reports. Many issuers also published quarterly results. Consequently, annual financial statements (unlike narratives) could be viewed as stale news, which should be irrelevant to the price formation process in informationally efficient markets. Perhaps this is also the reason why increases in sales and leverage are invariably statistically insignificant in our model specifications.

The *F*-statistics for the regressions indicate that the factors considered are jointly important from a statistical point of view. However, this observation must be tempered by the fact that the R-squared coefficients are relatively low. It is a well-established empirical finding that stock markets are excessively volatile compared to underlying fundamentals. Shiller (1981)

[5] Detailed results can be obtained from the authors upon request.

argues that price volatility is about five to thirteen times higher than that justified by a dividend-based valuation model. Returns are often a manifestation of the fickle sentiments of the investing public and can be orthogonal to the economic performance of companies. As a result, the impact of genuine fundamental drivers is intertwined with and obscured by noise, resulting in low values of the goodness of fit measures. This problem transcends our study and is a more general issue that has troubled financial economists since time immemorial.

7. Further Considerations

One possible extension of this study could be to investigate the market impact of words with negative connotations. To probe this issue, we use Henry's (2008:387) negativity thesaurus incorporating 85 words and measure the frequency with which these words appear in the qualitative parts of annual reports. We discover that most of our market reaction measures are negatively correlated with this frequency; however, these correlations are statistically insignificant. This clearly shows that positive words have greater explanatory power.

Perhaps investors do not believe that managers would voluntarily disclose bad news in narrative text, unless forced to do so by regulators. Managers are often remunerated by share or call option compensation schemes, which gives them incentives to suppress unfavourable information. These motivations are likely to be weakened during option granting periods when the desire to negotiate the lowest strike on the calls may dominate. However, it is safe to assume that managerial compensation increases with the stock price during the majority of periods. Due to the existence of this incentive, it is likely that the influence of bad news would be more apparent in financial statements rather than in the narrative. Davis and Tama-Sweet (2012) argue that managers tend to publish pessimistic narratives in outlets with the lowest impact and it is doubtful whether an annual report is such an outlet.

We are also not the first to claim that positive statements in UK annual reports are more informative compared to negative ones. Schleicher and Walker (2010), who examined the outlook section in chairmen's statements in conjunction with managerial incentives, found that tone is biased primarily by manipulating the number of negative statements (p. 388). Therefore the fact that negative words have less impact on the market attests to investors' rationality. There are also some parallels to the world of politics – Wisniewski and Moro (2014) find that policy makers are unlikely to draft pessimistic communiqués related to meetings in which they themselves have participated. As a result of these considerations, we believe that tone is more accurately measured by different shades of optimism rather than pessimism.

On an intuitive level, some may argue that positive words may be overrated, since managers have the proclivity to get involved in 'sugar-coating rituals'. However, rational investors are expected to fully anticipate such behaviour. Whenever managers fail to engage in such rituals, it may be a very strong signal to shareholders that there may be serious problems on the horizon. Consequently, this can give rise to a strong correlation between the degree of positiveness and market reaction.

Notwithstanding some of the aforementioned problems, we proceeded to construct an index which aggregates the frequencies of both positive and negative words. The practical issue is that the two thesauruses considered have different word counts, which unavoidably leads to the measured positive and negative word frequencies having different sample means and standard deviations. In order to make sure that both carry the same weight in the aggregation, we convert these frequencies into *z*-scores.[6] Consequently, we construct the following index:

$$Tone_i = Positivity\_z_i - Negativity\_z_i$$

[6] As a matter of convention, Diction software converts frequencies into *z*-scores before aggregating them.

Subsequently, we use the *Tone* variable in our market impact regressions instead of the *Positivity* indicator. The results presented in Table 6 reassure us that the statistical significance of the tone conveyed by a narrative is maintained, regardless of how this sentiment is defined.

[Table 6 about here]

Another potential problem is that our sample incorporates both a period of expansion and a subsequent recession sparked by the banking crisis. In order to control for this fact, we have constructed a dummy variable which takes a value of 1 if the annual report was published after 15th September 2008 (collapse of Lehman Brothers) and a value of 0 otherwise. This variable, however, proved insignificant in our regressions and does not alter the strength of the relationship between *Positiveness* and *CAR*s. Furthermore, we note that the correlations between different measures of market reaction and upbeat tone are always positive, irrespective of whether we look at the boom or bust sub-sample.

8. Conclusions

Our results indicate that the positiveness inherent in qualitative parts of annual reports, has a statistically significant association with abnormal returns around disclosure dates. More specifically, an upbeat tone typically induces statistically significant stock price increases. These results, that join a growing body of empirical evidence about the impact of narrative-based elements on markets, call for further and more detailed examination of qualitative parts of annual reports by both academics and practitioners. In particular, our findings affirm the usefulness of text-analysis software in revealing hidden characteristics of texts and thus suggest that such software tools may be fruitfully employed by investors and regulators alike. Computerized computational linguistic approaches to analyzing annual report narratives can be particularly helpful considering how voluminous these documents are. Previous studies focused

on shorter items, however reliable assessment of linguistic style in short documents is a rather problematic undertaking.

There are several facts that can be gleaned from our empirical observations. Firstly, although many claim that annual report narratives may have the tendency to suffer from subjective optimism, investors clearly believe that they also convey material information. In fact, they seem to rebalance their portfolios in response to the tone of the qualitative part of annual reports, which becomes apparent when examining the distribution of returns. While narratives may be partially used to build brands and manage impressions, they also appear to contribute to the reduction of informational asymmetries. Secondly, our study invites a reflection on the extent to which managers can exaggerate an optimistic message under the principles-based system operating in the UK. Most of the research on qualitative corporate outputs used US documents, which are produced under a rules-based regulatory framework. Our results, which record market reactions similar to those observed in the US, point to a similar set of general phenomena. We can therefore conclude that what restrains managers from injecting excessive positiveness bias into narratives is not only the litigation risk, but also potential reputational loss or other non-regulatory factors. The third lesson that can be drawn from our findings is that a thorough perusal of the narrative should be recommended. Whilst a number of previous studies performed content analysis of specific sections of annual reports, such as chairmen's statements, in our view it would be imprudent to advise investors to read parts of the annual report narratives selectively. This is not to say that some parts cannot contain more informational content than others, however, deliberately dismissing selected sections may be a misguided strategy. Fourthly, it appears that resources committed to drafting these documents are well spent. Companies typically involve many departments, accountants, lawyers, directors and external agencies to carefully design the message they wish to convey in their annual reports. This message is heard by market participants, who act accordingly. Lastly, we can infer

that the use of a semantic software package could, at least to a certain extent, be useful in predicting market reactions to annual report disclosures.

There are many avenues that further research could explore. Our computerized approach to text analysis is analytically elegant and convenient, however it is unable to assess the veracity of statements made or to evaluate whether managers are playing strategic disclosure games with investors and regulators. Further research needs to be conducted to answer these questions, which are outside the scope of the current paper. Secondly, the algorithm to measure positiveness employed here relies on computing frequencies from a user-specified thesaurus in the text. As such, it does not recognize sentence structures, subjunctive clauses or the context in which a given word occurs, even though all of these can modify or even negate the meaning of a particular word. Future research should endeavour to address these methodological deficiencies. Thirdly, we discover that narrative positiveness significantly correlates with announcement cumulative abnormal returns that measure the overall market response. However, this particular indicator is able to explain only a small proportion of the return variance. Inability to model price increases precisely is a well-known problem in finance, after Shiller (1981) pointed out that stock prices are substantially more volatile than underlying fundamentals. It is quite possible that much of the return variation is driven by non-fundamental, irrational factors which are difficult to capture in an empirical model. Our study points to only one incremental variable that may be useful to further explicate stock price fluctuations.

References

Abrahamson, E., Amir, E. (1996). The information content of the president's letter to shareholders. Journal of Business Finance & Accounting, 23(8), 1157-1182.

Ahern, K.R., Sosyura, D. (2014). Who writes the news? Corporate press releases during merger negotiations. Journal of Finance, 69(1), 241-291.

Ajzen, I., Fishbein, M. (2000). Attitudes and the attitude-behaviour relation: reasoned and automatic processes. European Review of Social Psychology, 11(1), 1-33.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. Journal of Financial Markets, 5(1), 31-56.

Anilowski, C., Feng, M., Skinner, D.J. (2007). Does earnings guidance affect market returns? The nature and information content of aggregate earnings guidance. Journal of Accounting and Economics, 44(1), 36-63.

Arbel, A., Carvell, S., Strebel, P. (1983). Giraffes, institutions and neglected firms. Financial Analysts Journal, 39(3), 57-63.

Atiase, R.K. (1985). Predisclosure information, firm capitalization, and security price behavior around earnings announcements. Journal of Accounting Research, 23(1), 21-36.

Baillie, W.M. (1974). Authorship attribution in Jacobean dramatic texts. In J.L. Mitchell (Ed.), Computers in the Humanities (pp. 73-81), Edinburgh: Edinburgh University Press.

Barbee, W.C., Mukherji, S. Jr., Raines, G.A. (1996). Do sales-price and debt-equity explain stock returns better than book-market and firm size? Financial Analysts Journal, 52(2), 56-60.

Ball, R., Brown, P. (1968). An empirical evaluation of accounting income numbers. Journal of Accounting Research, 6(2), 159-78.

Banz, R.W. (1981). The relationship between return and market value of common stocks. Journal of Financial Economics, 9(1), 3-18.

Bartov, E., Givoly, D., Hayn, C. (2002). The rewards to meeting or beating earnings expectations. Journal of Accounting and Economics, 33(2), 173-204.

Bhandari, L.C. (1988). Debt/equity ratio and expected common stock returns: empirical evidence. Journal of Finance, 43(2), 507-528.

Bhattacharya, U., Galpin, N., Ray, R., Yu, X. (2009). The role of the media in the internet IPO bubble. Journal of Financial and Quantitative Analysis, 44(2), 657-682.

Brennan, N.M., Guillamon-Saorin, E., Pierce, A. (2009). Methodological insights. Accounting, Auditing & Accountability Journal, 22(5), 789-832.

Brown, L.D., Caylor, M.L. (2005). A temporal analysis of quarterly earnings thresholds: propensities and valuation consequences. Accounting Review, 80(2), 423-440.

Brown, S.J., Warner, J.B. (1980). Measuring security price performance. Journal of Financial Economics, 8(3), 205-258.

Brown, S.J., Warner, J.B. (1985). Using daily stock returns: the case of event studies. Journal of Financial Economics, 14(1), 3-31.

Campbell, J.Y., Lo, A.W., MacKinley, A.C. (1997). The Econometrics of Financial Markets. Princeton: Princeton University Press.

Chatterjee, S., Price, B. (1991). Regression Diagnostics. New York: John Wiley.

Cho, C.H., Roertd, R.W., Patten, D.M. (2010). The language of us corporate environmental disclosure. Accounting, Organizations and Society, 35(4), 431-443.

Cicon, J., Clarke, J., Ferris, P.F., Jayaraman, N. (2014). Managerial expectations of synergy and the performance of acquiring firms: the contribution of soft data. Journal of Behavioral Finance, 15(3), 161-174.

Clatworthy, M., Jones, M.J. (2003). Financial reporting of good news and bad news: evidence from accounting narratives. Accounting and Business Research, 33(3), 171-185.

Craig, R.J., Brennan. N.M. (2012). An exploration of the relationship between language choice in CEO letters to shareholders and corporate reputation. Accounting Forum, 36(3), 166-177.

Crano, W.D., Prislin, R. (2006). Attitudes and persuasion. Annual Review of Psychology, 57, 345-374.

Davies, P. (2007). Davies review of issuer liability: liability for misstatements to the market, available on the internet at http://webarchive.nationalarchives.gov.uk/20100407010852/http:/www.hm-treasury.gov.uk/d/davies_discussionpaper_260307.pdf, Accessed 28.07.2015.

Davies, P. (2009). Liability for misstatements in the market: some reflections. Journal of Corporate Law Studies, 9(2), 295-313.

Davis, A.K., Ge, W., Matsumoto, D., Zhang, J.L. (2015). The effect of manager-specific optimism on the tone of earnings conference calls. Review of Accounting Studies, 20(2), 639-673.

Davis, A.K., Piger, J.M., Sedor, L.M. (2012). Beyond the numbers: measuring the information content of earnings press release language. Contemporary Accounting Research, 29(3), 845-868.

Davis, A.K., Tama-Sweet, I. (2012). Managers' use of language across alternative disclosure outlets: earnings press releases versus MD&A. Contemporary Accounting Research, 29(3), 804-837.

DeFond, M.L., Park, C.W. (2001). The reversal of abnormal accruals and the market valuation of earnings surprises. Accounting Review, 76(3), 375-404.

Demers, E., Vega, C. (2010). Soft Information in Earnings Announcements: News or Noise? (INSEAD Faculty & Research Working Paper 2010/33/AC), available on the internet at http://www.insead.edu/facultyresearch/faculty/personal/edemers/documents/SoftInformationinEarningsAnnouncementsNewsorNoise-INSEADWP.pdf Accessed 08.07.15.

Doran, J.S., Peterson, D.R., Price, S.M. (2012). Earnings conference call content and stock price: the case of REITs. Journal of Real Estate Finance and Economics, 45(2), 402-434.

Easton, P.D., Zmijewski, M.E. (1989). Cross-sectional variation in the stock market response to accounting earnings announcements. Journal of Accounting and Economics, 11(2-3), 117-141.

Fama, E.F. (1970). Efficient capital markets: a review of theory and empirical work. Journal of Finance, 25(2), 383-417.

Fama, E.F., French, K.R. (1992). The cross-section of expected returns. Journal of Finance, 47(2), 427-465.

Fama, E.F., French, K.R. (1993). Common risk factors in the returns on stocks and bonds. Journal of Financial Economics, 33(1), 3-56.

Feldman, R., Govindaraj, S., Livnat, J., Segal, B. (2008). The Incremental Information Content of Tone Change in Management Discussion and Analysis (SSRN Working Paper), available on the internet at http://ssrn.com/abstract=1126962 Accessed 08.07.15.

Ferguson, N.J., Philip, D., Lam, H.Y.T., Guo, J.M. (2015). Media content and stock returns: the predictive power of press. Multinational Finance Journal, 19(1), 1-31.

Ferran, E. (2009). Are the US-style investor suits coming to the UK? Journal of Corporate Finance Law Studies, 9(2), 315-348.

Francis, J., Schipper, K., Vincent, L. (2002). Disclosures and the increased usefulness of earnings announcements. Accounting Review, 77(3), 515-546.

Frankel, R., Mayew, W., Sun, Y. (2010). Do pennies matter? Investor relations consequences of small negative earnings surprises. Review of Accounting Studies, 15(1), 220-242.

Frazier, K.B., Ingram, R.W., Tennyson, B.M. (1984). A methodology for the analysis of narrative accounting disclosures. Journal of Accounting Research, 22(1), 318-331.

Freeman, R.N. (1987). The association between accounting earnings and security returns for large and small firms. Journal of Accounting and Economics, 9(2), 195-228.

Frost, C.A., Pownall, G. (1994). Accounting disclosure practices in the United States and the United Kingdom. Journal of Accounting Research, 32(1), 75-102.

Gilbert, A., Tourani-Rad, A., Wisniewski, T.P. (2006). Do insiders crowd out analysts? Finance Research Letters, 3(1), 40-48.

Gordon, E.A., Henry, E., Peytcheva, M., Sun, L. (2013). Discretionary disclosure and the market reaction to restatements. Review of Quantitative Finance and Accounting, 41(1), 75-110.

Grimmer, J., Stewart, B. M. (2013). Text as data: the promise and pitfalls of automatic content analysis methods for political texts. Political Analysis, 21(3), 267-297.

Henry, E. (2006). Market reaction to verbal components of earnings press releases: event study using a predictive algorithm. Journal of Emerging Technologies in Accounting, 3, 1-19.

Henry, E. (2008). Are investors influenced by how earnings press releases are written? Journal of Business Communication, 45(4), 363-407.

Henry, E., Leone, A.J. (2009). Measuring Qualitative Information in Capital Markets Research (SSRN Working Paper), available on the internet at http://ssrn.com/abstract=1470807 Accessed 08.07.15.

Hildebrandt, H.W., Snyder, R.D. (1981). The Pollyanna hypothesis in business writing: initial results, suggestions for research. International Journal of Business Communication, 18(1), 5-15.

Hoskin, R., Hughes, J., Ricks, W. (1986). Evidence on the incremental information content of additional firm disclosures made concurrently with earnings. Journal of Accounting Research, 24(Supplement), 1–32.

Huang, X., Teoh, S.H., Zhang, Y. (2014). Tone management. Accounting Review, 89(3), 1083-1113.

Hughes, J.S., Ricks, W.E. (1987). Associations between forecast errors and excess returns near to earnings announcement. Accounting Review, 62(1), 158-175.

Jordan, C.E., Waldron, M.A., Clark, S.J. (2007). An analysis of the comparative predictive abilities of operating cash flows, earnings, and sales. Journal of Applied Business Research, 23(3), 53-60.

Kothari, S.P., Li, X., Short, J.E. (2009). The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: a study using content analysis. Accounting Review, 84(5), 1639-1670.

Kothari, S.P., Short, J. (2003). The Effect of Disclosures by Management, Analysts, and Financial Press on the Equity Cost of Capital, available on the internet at: http://ebusiness.mit.edu/research/papers/195__shortkothari_disclosures.pdf Accessed 15.12.2014.

Ledger, G., Merriam, T. (1994). Shakespeare, Fletcher, and the two noble kinsmen. Literary and Linguistic Computing, 9(3), 235-248.

Lesmond, D.A, Ogden, J.P., Trzcinka, C.A. (1999). A new estimate of transaction costs. Review of Financial Studies, 12(5), 1113-1141.

Lev, B. (1989). On the usefulness of earnings and earnings research: lessons and directions from two decades of empirical research. Journal of Accounting Research, 27, 153-192.

Li, F. (2010). The information content of forward-looking statements in corporate filings – a naïve machine learning approach. Journal of Accounting Research, 48(5), 1049-1102.

Loughran, T., McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. Journal of Finance, 66(1), 35-65.

Malaviya, P., Meyers-Levy, J., Sternthal, B. (1999). Ad repetition in a cluttered environment: the influence of type of processing. Psychology & Marketing, 16(2), 99-118.

McConnell, D., Haslem, J.A. & Gibson, V.R. (1986). The president's letter to shareholders: a new look. Financial Analysts Journal, 42(5), Sep-Oct, 66-70.

Merkl-Davies, D.M., Brennan, N.M., McLeay, S.J. (2011). Impression management and retrospective sense-making in corporate narratives: a social psychology perspective. Accounting, Auditing & Accountability Journal, 24(3), 315-344.

Merkley, K.J. (2014). Narrative disclosure and earnings performance: evidence from R&D disclosures. Accounting Review, 89(2), 725-757.

Nelson, M.W. (2003). Behavioral evidence on the effects of principles- and rules-based standards. Accounting Horizons, 17(1), 91-104.

Parhankangas, A., Ehrlich, M. (2014). How entrepreneurs seduce business angles: an impression management approach. Journal of Business Venturing, 29(4), 543-564.

Perloff, R.M. (2010). The Dynamics of Persuasion: Communication and Attitudes in the Twenty-First Century. New York: Routledge.

Petty, R.E., Cacioppo, J.T. (1986). The elaboration likelihood model of persuasion. Advances in Experimental Social Psychology, 19, 123-205.

Petty, R.E., Cacioppo, J.T., Schumann, D. (1983). Central and peripheral routes to advertising effectiveness: the moderating role of involvement. Journal of Consumer Research, 10(2), 135-146.

Petty, R., Cacioppo, J.T. (2011). Communication and Persuasion: Central and Peripheral Routes to Attitude Change. New York: Springer-Verlag.

Price, S., Doran, J., Peterson, D., Bliss, B. (2012). Earnings conference calls and stock returns: the incremental informativeness of textual tone. Journal of Banking and Finance, 36(4), 992-1011.

Rogers, J.L., Van Buskirk, A., Zechman, S.L.C. (2011). Disclosure tone and shareholder litigation. Accounting Review, 86(6), 2155-2183.

Rosenberg, B., Reid, K., Lanstein, R. (1985). Persuasive evidence of market inefficiency. Journal of Portfolio Management, 11(3), 9-16.

Rutherford, B.A. (2005). A corpus linguistics-based approach. International Journal of Business Communication, 42(4), 349-378.

Scharnd, C.M., Walther, B.R. (2000). Strategic benchmarks in earnings announcements: the selective disclosure of prior-period earnings components. Accounting Review, 75(2), 151-177.

Schleicher, T., Walker, M. (2010). Bias in the tone of forward-looking narratives. Accounting and Business Research, 40(4), 371-390.

SEC (1998). A plain English handbook: how to create clear SEC disclosure documents, available on the internet at http://www.sec.gov/pdf/handbook.pdf. Accessed 28.07.15.

SEC (2015). Year-by-year SEC enforcement statistics. Available on the internet at https://www.sec.gov/news/newsroom/images/enfstats.pdf, Accessed 28.07.15.

Shiller, R. J. (1981). Do stock prices move too much to be justified by subsequent changes in dividends? American Economic Review, 76(3), 483–98.

Shumway, T. (2001). Forecasting bankruptcy more accurately: a simple hazard model. Journal of Business, 74(1), 101-124.

Shin, H.S. (1994). News management and the value of firms. RAND Journal of Economics, 25(1), 58-71.

Smith, M., Taffler, R.J. (1992). The chairman's statement and corporate financial performance. Accounting and Finance, 32(2), 75-90.

Smith, M., Taffler, R.J. (1995). The incremental effect of narrative accounting information in corporate annual reports. Journal of Business Finance & Accounting, 22(8), 1195-1210.

Smith, M., Taffler, R.J. (2000). The chairman's statement: a content analysis of discretionary narrative disclosures. Accounting, Auditing & Accountability Journal, 13(5), 624-646.

Sponholtz, C. (2008). The information content of earnings announcements in Denmark. International Journal of Managerial Finance, 4(1), 4-36.

Sussman, S.W., Siegal, W.S. (2003). Informational influence in organizations: an integrated approach to knowledge adoption. Information Systems Research, 14(1), 47-65.

Swales, G.S. Jr. (1988). Another look at the president's letter to stockholders. Financial Analysts Journal, 44(2), Mar.–Apr., 71-73.

Sydserff, R., Weetman, P. (2002). Developments in content analysis: a transitivity index and scores. Accounting, Auditing & Accountability Journal, 15(4), 523-545.

Tetlock, P.C. (2007). Giving content to investor sentiment: the role of media in the stock market. Journal of Finance, 62(3), 1139-68.

Tetlock, P.C., Saar-Tsechansky, M., Macskassy, S. (2008). More than words: quantifying language to measure firms' fundamentals. Journal of Finance, 63(3), 1437-67.

Weiss, R.F. (1969). Repetition of Persuasion. Psychological Reports, 25(2), 669-670.

Wilson, W., Miller, H. (1968). Repetition, order of presentation, and timing of arguments and measures as determinants of opinion change. Journal of Personality and Social Psychology, 9(2), 184-188.

Wisniewski, T.P. (2004). Reexamination of the link between insider trading and price efficiency. Economic Systems, 28(2), 209-228.

Wisniewski, T.P., Moro, A. (2014). When the leaders speak, the markets listen. European Accounting Review, 23(4), 519-551.

Yuthas, K., Rogers, R., Dillard, J.F. (2002). Communicative action and corporate annual reports. Journal of Business Ethics, 41(1-2), 141-157.

Table 1
Definitions of Explanatory Variables

Panel A. Explanatory Variables Used in the 'Empirical Results' Section	
Variable	**Definition**
Positiveness	Frequency with which the positive words listed in Henry's (2008) tone thesaurus appear in the narrative of the annual report. For the complete list of these words see Appendix I. The frequency is defined as the number of positive words divided by the number of total words in the document
Size	Natural logarithm of company's capitalization at the end of fiscal year to which the annual report refers
Book_to_Market	Ratio of book value per share to share price at the end of fiscal year covered by the report
Earnings_Surprise	Increase in earnings per share from the previous year scaled by the share price measured at the end of fiscal year
Δ%Sales	Percentage increase in sales relative to the previous year
ΔLeverage	Increase in leverage, where leverage is defined as total liabilities over total assets
Panel B. Explanatory Variable Used in the 'Further Considerations' Section	
Variable	**Definition**
Tone	To construct this variable, we measured the frequencies with which the positive and negative words listed in Henry's (2008) tone thesauruses appear in the narrative of the annual reports. These frequencies have been subsequently converted into *z*-scores by deducting their individual sample means and dividing by standard deviation. *Tone* for a particular report is measured as its positive word frequency *z*-score minus the negative word frequency *z*-score

Table 2
Summary Statistics

Variable	Mean	Standard Deviation	25th Percentile	Median	75th Percentile
CAR_MA(-1,1)	0.0671%	3.9994%	-1.7751%	0.0086%	1.9135%
CAR_MA(-1,5)	0.2502%	5.9793%	-2.5880%	-0.0988%	2.8867%
CAR_MM(-1,1)	0.0207%	3.5317%	-1.5615%	-0.0795%	1.4735%
CAR_MM(-1,5)	0.0450%	5.2468%	-2.4810%	-0.1416%	2.3364%
Positiveness	0.4778%***	0.4071%	0.2180%	0.3960%	0.5960%
Size	14.3314***	1.4504	13.3439	14.0619	15.0463
Book_to_Market	0.5897***	0.7472	0.2323	0.4024	0.7566
Earnings_Surprise	-0.0168	0.4404	-0.0115	0.0076	0.0258
Δ%Sales	8.3288%***	26.1616%	0.3921%	7.3865%	15.6980%
ΔLeverage	-0.0046*	0.0979	-0.0401	-0.0046	0.0280

Note: CAR_MA denotes cumulative abnormal return from a constant-mean-adjusted model and the parameters in the parentheses denote the length of the event window. CAR_MM are cumulative returns in excess of a market model benchmark. The remaining variables are defined in Table 1. A two-tailed test for the hypothesis that the mean of a variable is equal to zero has been performed. *, **, *** reported in the 'Mean' column denote rejection at 90%, 95% and 99% confidence level, respectively.

Table 3
Pearson Correlation Table

	CAR_MA (-1,1)	CAR_MA (-1,5)	CAR_MM (-1,1)	CAR_MM (-1,5)	Positiveness	Size	Book_to_ Market	Earnings Surprise	Δ%Sales	ΔLeverage
CAR_MA(-1,1)	1.0000									
CAR_MA(-1,5)	0.6645***	1.0000								
CAR_MM(-1,1)	0.8761***	0.5806***	1.0000							
CAR_MM(-1,5)	0.5947***	0.8744***	0.6780***	1.0000						
Positiveness	0.0547**	0.0515*	0.0647**	0.0771***	1.0000					
Size	-0.0467*	-0.0742***	-0.0316	-0.0550**	-0.0077	1.0000				
Book_to_Market	0.0348	0.0367	0.0089	-0.0050	-0.0120	-0.1378***	1.0000			
Earnings_Surprise	0.0265	-0.0498*	0.0145	-0.0502*	0.0220	0.0824***	-0.4344***	1.0000		
Δ%Sales	0.0204	0.0143	-0.0047	-0.0257	-0.0266	0.0080	-0.0342	0.0420	1.0000	
ΔLeverage	0.0149	0.0390	0.0075	0.0079	0.0237	-0.0302	0.0144	-0.1536***	-0.0497*	1.0000

Note: The first four variables in the table measure the cumulative abnormal returns computed using a constant-mean adjustment (MA) and market-model adjustment (MM). For definitions of the remaining variables please refer to Table 1. A two-tailed test for the hypothesis that the correlation coefficient is equal to zero has been performed. *, **, *** denote rejection at 90%, 95% and 99% confidence level, respectively.

Table 4

Determinants of Constant-Mean-Adjusted Cumulative Abnormal Returns

Panel A. Regressions on CAR_MA(-1,1)			
	(1)	(2)	(3)
Intercept	-0.0019	0.0139	0.0134
	(0.0016)	(0.0109)	(0.0111)
Positiveness	0.5286**	0.5278**	0.5341**
	(0.2614)	(0.2619)	(0.2660)
Size		-0.0012	-0.0012
		(0.0007)	(0.0008)
Book_to_Market		0.0017	0.0029*
		(0.0014)	(0.0016)
Earnings_Surprise			0.0050*
			(0.0028)
Δ%Sales			0.0035
			(0.0041)
ΔLeverage			0.0089
			(0.0115)
R-squared	0.2895%	0.6013%	0.9046%
F-stat	4.0880	2.8252	2.0966
Prob (F-stat)	0.0434	0.0375	0.0509
No. obs.	1410	1405	1385
Panel B. Regressions on CAR_MA(-1,5)			
	(1)	(2)	(3)
Intercept	-0.0013	0.0380	0.0385**
	(0.0025)	(0.0163)	(0.0165)
Positiveness	0.7865**	0.7756**	0.7640*
	(0.3909)	(0.3907)	(0.3963)
Size		-0.0028**	-0.0028**
		(0.0011)	(0.0011)
Book_to_Market		0.0021	0.0010
		(0.0022)	(0.0024)
Earnings_Surprise			-0.0049
			(0.0041)
Δ%Sales			0.0045
			(0.0062)
ΔLeverage			0.0195
			(0.0171)
R-squared	0.2868%	0.8689%	1.1473%
F-stat	4.0492	4.0933	2.6654
Prob (F-stat)	0.0444	0.0066	0.0142
No. obs.	1410	1405	1385

Note: This table reports regressions where the constant-mean-adjusted cumulative returns are taken to act as a dependent variable. Panel A models the CAR measured in the (-1,1) event window, while Panel B extends the window to (-1,5). All of the explanatory variables are defined in Table 1. The table presents coefficient estimates with the corresponding standard errors in parentheses, coefficient of determination, the F-test for the null hypothesis that the regressors are jointly statistically insignificant and the number of observations. *, **, *** denote statistical significance at 10%, 5% and 1%, respectively.

Table 5
Determinants of Market-Model-Adjusted Cumulative Abnormal Returns

Panel A. Regressions on CAR_MM(-1,1)			
	(1)	(2)	(3)
Intercept	-0.0024*	0.0082	0.0080
	(0.0014)	(0.0097)	(0.0098)
Positiveness	0.5456**	0.5438**	0.5586**
	(0.2307)	(0.2314)	(0.2353)
Size		-0.0008	-0.0008
		(0.0007)	(0.0007)
Book_to_Market		0.0004	0.0007
		(0.0013)	(0.0014)
Earnings_Surprise			0.0019
			(0.0024)
Δ%Sales			-0.0004
			(0.0037)
ΔLeverage			0.0031
			(0.0102)
R-squared	0.3955%	0.5007%	0.5655%
F-stat	5.5911	2.3501	1.3062
Prob (F-stat)	0.0182	0.0708	0.2511
No. obs.	1410	1405	1385
Panel B. Regressions on CAR_MM(-1,5)			
	(1)	(2)	(3)
Intercept	-0.0042*	0.0249*	0.0262*
	(0.0022)	(0.0143)	(0.0145)
Positiveness	0.9759***	0.9613***	1.0017***
	(0.3425)	(0.3428)	(0.3477)
Size		-0.0020**	-0.0020**
		(0.0010)	(0.0010)
Book_to_Market		-0.0010	-0.0029
		(0.0019)	(0.0021)
Earnings_Surprise			-0.0078**
			(0.0036)
Δ%Sales			-0.0045
			(0.0054)
ΔLeverage			-0.0034
			(0.0150)
R-squared	0.5733%	0.8603%	1.2954%
F-stat	8.1189	4.0527	3.0142
Prob (F-stat)	0.0044	0.0070	0.0062
No. obs.	1410	1405	1385

Note: This table reports the estimates of regressions where the market-model-adjusted cumulative abnormal return is the dependent variable. The results in Panel A refer to the CAR computed in the (-1,1) window, whereas Panel B is based on the (-1,5) window. Standard errors are given in parentheses below the parameter estimates. R-square, F-statistic for the joint significance of explanatory variables and the number of observations are presented at the bottom of each panel. *, **, *** denote statistical significance at 10%, 5% and 1%, respectively.

Table 6
Market Reaction and the Tone of Annual Report Narrative

	Determinants of Constant-Mean-Adjusted Returns		Determinants of Market-Model-Adjusted Returns	
	(1)	(2)	(3)	(4)
Intercept	0.0007	0.0164	0.0002	0.0112
	(0.0011)	(0.0110)	(0.0009)	(0.0097)
Tone	0.0013*	0.0015*	0.0015**	0.0016**
	(0.0008)	(0.0008)	(0.0007)	(0.0007)
Size		-0.0012		-0.0008
		(0.0008)		(0.0007)
Book_to_Market		0.0031*		0.0009
		(0.0016)		(0.0014)
Earnings_Surprise		0.0051*		0.0021
		(0.0028)		(0.0024)
Δ%Sales		0.0032		-0.0007
		(0.0041)		(0.0037)
ΔLeverage		0.0091		0.0033
		(0.0115)		(0.0102)
R-squared	0.2100%	0.8624%	0.3413%	0.5364%
F-stat	2.9630	1.9978	4.8213	1.2387
Prob (F-stat)	0.0854	0.0630	0.0283	0.2835
No. obs.	1410	1385	1410	1385

Note: This table reports the estimates of regressions where the abnormal returns in the (-1,1) event window act as the dependent variable. Standard errors are given in parentheses below the parameter estimates. R-square, F-statistic for the joint significance of explanatory variables and the number of observations are presented at the bottom of the table. *, **, *** denote statistical significance at 10%, 5% and 1%, respectively.

Appendix I

Thesaurus of Positive Words

positive, positives, success, successes, successful, succeed, succeeds, succeeding, succeeded, accomplish, accomplishes, accomplishing, accomplished, accomplishment, accomplishments, strong, strength, strengths, certain, certainty, definite, solid, excellent, good, leading, achieve, achieves, achieved, achieving, achievement, achievements, progress, progressing, deliver, delivers, delivered, delivering, leader, leading, pleased, reward, rewards, rewarding, rewarded, opportunity, opportunities, enjoy, enjoys, enjoying, enjoyed, encouraged, encouraging, up, increase, increases, increasing, increased, rise, rises, rising, rose, risen, improve, improves, improving, improved, improvement, improvements, strengthen, strengthens, strengthening, strengthened, stronger, strongest, better, best, more, most, above, record, high, higher, highest, greater, greatest, larger, largest, grow, grows, growing, grew, grown, growth, expand, expands, expanding, expanded, expansion, exceed, exceeds, exceeded, exceeding, beat, beats, beating

Appendix II

Companies and Number of Annual Reports Included in the Sample

3i Group	7	Capita	7
3i Infrastructure	5	Carillion	7
Admiral Group	7	Carnival	7
Amec	7	Carpetright	7
Amlin	7	Catlin Group	7
Anglo American	7	Centrica	7
Antofagasta	7	Close Brothers Group	7
Arm Holdings	7	Cobham	7
Ashmore Group	6	Colt Group	7
Ashtead Group	7	Compass Group	7
Associated Brit.Foods	7	Computacenter	7
Astrazeneca	7	CRH	7
Aveva Group	7	Croda International	7
Aviva	7	CSR	7
Babcock Intl.	7	Daejan Holdings	7
BAE Systems	7	Dairy Crest	7
Balfour Beatty	7	De La Rue	7
Barratt Developments	7	Debenhams	6
BBA Aviation	7	Dechra Pharmaceuticals	7
Beazley	7	Diageo	6
Bellway	7	Dialight	7
Berendsen	7	Diploma	6
Berkeley Group Hdg.(The)	7	Dixons Retail	7
BG Group	7	Domino Printing Sciences	7
BHP Billiton	7	Drax Group	7
Big Yellow Group	7	Dunelm Group	6
Blackrock World Mng.	5	Electrocomp.	7
Bodycote	7	Elementis	7
Booker Group	6	Eurasian Natres.Corp.	5
Bovis Homes Group	7	Experian	5
BP	7	Fenner	7
Brewin Dolphin	7	Ferrexpo	6
British American Tobacco	7	Fidessa Group	7
British Land	7	First Group	7
British Sky Bcast.Group	7	Fresnillo	5
Britvic	7	G4S	6
BT Group	7	Galliford Try	7
BTG	7	Genus	7
Bunzl	7	GKN	7
Burberry Group	7	Glaxosmithkline	7
Bwin Party Digital Entm.	7	Glencore Xstrata	2
Cable & Wireless Comms.	7	Go-Ahead Group	6

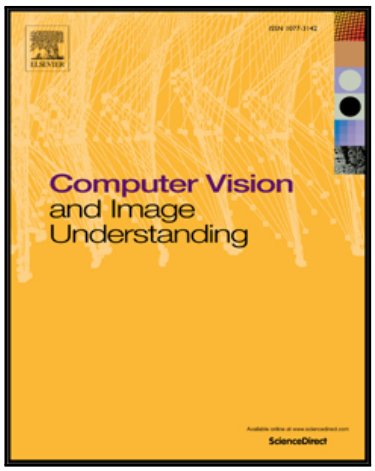
Companies and Number of Annual Reports Included in the Sample (Continued)

Grainger	7	Millennium & Cpth.Htls.	7
Great Portland Estates	7	Mitchells & Butlers	7
Greene King	7	MITIE Group	7
Greggs	7	Mondi	5
Halfords Group	7	National Express	7
Halma	7	National Grid	7
Hammerson	7	Next	7
Hansteen Holdings	7	Oxford Instruments	7
Hargreaves Lansdown	5	Paragon Gp.Of Cos.	7
Hays	7	Paypoint	6
Henderson Group	7	Pearson	7
HICL Infrastructure	6	Pennon Group	7
Hikma Pharmaceuticals	7	Persimmon	6
Hiscox	7	Petrofac	7
Hochschild Mining	6	Phoenix Group Hdg. (Lon)	3
Home Retail Group	6	Polar Capital Tech.Tst.	7
Homeserve	7	Premier Farnell	7
Howden Joinery Gp.	7	PZ Cussons	7
Hunting	7	Qinetiq Group	6
ICAP	7	Randgold Resources	7
Ictl.Htls.Gp.	7	Rank Group	6
IG Group Holdings	7	Rathbone Brothers	7
IMI	7	Reckitt Benckiser Group	7
Imperial Tobacco Gp.	7	Redrow	7
Inchcape	7	Reed Elsevier	7
Informa	7	Regus	7
Inmarsat	7	Renishaw	7
Intermediate Capital Gp.	7	Rentokil Initial	7
Interserve	7	Resolution	4
Intertek Group	7	Restaurant Group	7
Invensys	7	Rexam	7
IP Group	7	Rightmove	7
ITE Group	7	Rio Tinto	7
ITV	6	RIT Capital Partners	7
Jardine Lloyd Thompson	7	Rotork	7
Kazakhmys	7	RPC Group	7
Kcom Group	7	RPS Group	7
Kenmare Res. (Lon)	7	Sabmiller	7
Kier Group	7	Sage Group	7
Kingfisher	7	Salamander Energy	6
Ladbrokes	7	Savills	7
Lonmin	7	Schroders	7
Man Group	7	Segro	7
Marks & Spencer Group	7	Senior	7
Meggitt	7	Serco Group	7
Menzies (John)	7	Severn Trent	7
Michael Page Intl.	7	Shaftesbury	7

Companies and Number of Annual Reports Included in the Sample (Continued)

SIG	7	Tesco	7
Smith & Nephew	7	Travis Perkins	7
Smiths Group	7	Tullett Prebon	6
SOCO International	7	Tullow Oil	7
Spectris	7	UBM	7
Spirax-Sarco	7	Ultra Electronics Hdg.	7
Spirent Communications	7	Unilever (UK)	7
Sports Direct Intl.	5	Vedanta Resources	7
SSE	7	Victrex	7
SVG Capital	7	Vodafone Group	7
Synergy Health	7	Weir Group	7
Tate & Lyle	5	WH Smith	7
Taylor Wimpey	6	Whitbread	7
Ted Baker	6	William Hill	7
Telecity Group	5	WPP	7
Telecom Plus	7		

Note: This table presents a list of companies included in the sample and the number of annual reports used for each of the companies. The number of reports sums to 1410, which is our sample size.



Hierarchical Transfer Learning for Online Recognition of Compound Actions

Victoria Bloom , Vasileios Argyriou , Dimitrios Makris

PII: S1077-3142(15)00264-7
DOI: 10.1016/j.cviu.2015.12.001
Reference: YCVIU 2355

To appear in: *Computer Vision and Image Understanding*

Received date: 21 December 2014
Revised date: 26 November 2015
Accepted date: 3 December 2015

Please cite this article as: Victoria Bloom , Vasileios Argyriou , Dimitrios Makris , Hierarchical Transfer Learning for Online Recognition of Compound Actions, *Computer Vision and Image Understanding* (2015), doi: 10.1016/j.cviu.2015.12.001

Highlights

- A novel online action recognition method for fast detection of compound actions.
- A key contribution is a transfer learning strategy from simple to complex datasets.
- Another key contribution is an automatically configured hierarchical body model.
- Experimental results show an improvement in action recognition performance of 16%.
- The proposed algorithm is real-time with an average latency of just 2 frames.

Hierarchical Transfer Learning for Online Recognition of Compound Actions

Victoria Bloom [a,b,*], Vasileios Argyriou [a], Dimitrios Makris [a]

[a] Digital Imaging Research Centre, Kingston University, United Kingdom

[b] Coventry University, United Kingdom

Abstract

Recognising human actions in real-time can provide users with a natural user interface (NUI) enabling a range of innovative and immersive applications. A NUI application should not restrict users' movements; it should allow users to transition between actions in quick succession, which we term as compound actions. However, the majority of action recognition researchers have focused on individual actions, so their approaches are limited to recognising single actions or multiple actions that are temporally separated.

This paper proposes a novel online action recognition method for fast detection of compound actions. A key contribution is our hierarchical body model that can be automatically configured to detect actions based on the low level body parts that are the most discriminative for a particular action. Another key contribution is a transfer learning strategy to allow the tasks of action segmentation and whole body modelling to be performed on a related but simpler dataset, combined with automatic hierarchical body model adaption on a more complex target dataset.

Experimental results on a challenging and realistic dataset show an improvement in action recognition performance of 16% due to the introduction of our hierarchical transfer learning. The proposed algorithm is fast with an average latency of just 2 frames (66ms) and outperforms state of the art action recognition algorithms that are capable of fast online action recognition.

*Corresponding author.
E-mail addresses: Victoria.Bloom@kingston.ac.uk, Victoria.Bloom@coventry.ac.uk (V. Bloom), Vasileios.Argyriou@kingston.ac.uk (V. Argyriou), D.Makris@kingston.ac.uk (D. Makris).

1. Introduction

The research field of human action recognition has rapidly expanded in recent years with many innovative applications in a range of sectors including healthcare, education and entertainment. In healthcare, action recognition enables touch-free browsing of medical images in operating rooms, physical therapy at home and in clinics and for patient monitoring. In education, action recognition can increase the engagement of users by providing realistic and immersive training simulations. In entertainment, action recognition enables touch-free interaction with Smart TVs and games consoles for more intuitive and natural interaction. A key requirement of these interactive applications is the ability to robustly detect actions in real-time so the system can provide an appropriate response to the user with no apparent delay.

Historically, action recognition research has focused on increasing accuracy on datasets in highly controlled environments. These datasets normally contained a single person that was instructed to perform a single action clearly (see Figure 1). Recognition was performed offline after viewing a complete sequence and algorithms were evaluated by the number of correctly classified sequences. A recent survey [1] showed perfect or near perfect action recognition accuracy on simple datasets with a small number of actions.

Figure 1 Simple boxing sequence with a single person performing a punch (KTH) [3]

The traditional offline approach led to simplification of the problem, overinflated accuracy and lack of applicability to real world situations. Recent research toward more realistic action recognition has changed to online action recognition where different actions are detected in real-time whilst they are being observed. However, the focus has been on recognising actions which are temporally well separated and easy to segment. In contrast, this work considers multiple actions performed in quick succession, which are critical for robust

action detection in natural user interface (NUI) applications. When multiple actions are performed in quick succession movements from different actions may temporally overlap resulting in complex poses, which we term as compound actions. For example, in a full body fighting game a player may throw punches in quick succession, one arm may still be finishing the previous punch whilst the other arm is performing the next punch or a player may leave one arm in the defend position and punch with the other arm (as shown in Figure 2). Detecting multiple actions in quick succession is a more complex problem than recognising actions which are temporally well separated.

Figure 2 Complex fighting sequences between multiple players, performing multiple actions in quick succession so that the movements temporally overlap (G3Di) [4]. Each row represents a different sequence with visual examples taken every 3 frames.

Existing work on recognising more complex actions has to date only been researched in an offline context. To evaluate the performance of action recognition algorithms on more realistic actions several datasets have been extracted from TV and film (YouTube Action Dataset [5], Hollywood Human Actions Dataset [6], UCF sports action dataset [7]). In these datasets the actions are performed in real-world scenarios with diverse and cluttered backgrounds as well as significant changes in viewpoint. The individual actions are realistic but the major limitation of these datasets is that they have been segmented into sequences containing a single action suitable for offline action recognition. The diversity and complexity of real-world datasets makes accurate labelling difficult and time consuming. To overcome this problem Ma et al. [8] employed transfer learning to transfer knowledge from a simpler domain (e.g. KTH [3]) to a more complex target domain (e.g. YouTube Action Dataset) but their approach was limited to offline action recognition. An area that has not been considered before is the potential for transfer learning to improve online action recognition.

Several NUI datasets with multiple actions in each sequence have been captured (MSRC-12 [9], G3D [10], G3Di [4]) and action points [11] provided, as temporal anchors to enable evaluation of online action recognition algorithms. Good performance has been achieved on the datasets where the actions were recorded under controlled circumstances (MSRC-12, G3D) but performance dramatically decreased when the same algorithm [4] was applied to a real-world scenario of a full body fighting game (G3Di). All three datasets contain multiple actions but the difference is that the MSRC-12 and G3D datasets contain actions that are temporally well separated whereas the G3Di dataset, contains transitions between actions and even multiple actions at the same time. Temporal merging of a user's actions results in compound actions comprising of movements from different actions, which have not been adequately addressed by existing approaches.

In this work we propose a novel hierarchical transfer learning algorithm for online action recognition of compound actions. Specifically, transfer learning is employed to allow the tasks of action segmentation and modelling to be performed on a related but simpler dataset, combined with model adaptation to improve performance on a more complex dataset. Furthermore, we represent actions hierarchically to provide the flexibility to recognise poses that are not in

the source dataset by introducing independence between limbs. Evaluation on a realistic and challenging public action dataset confirms the effectiveness of our approach.

2. Literature Review

A key requirement of many real-world applications is the ability to recognise actions online. However, recent surveys [12], [13] show that the majority of existing action recognition algorithms are offline and rely on observing a pre-segmented action sequence before classification of a single action. A common adaptation of existing approaches is to use a sliding window and classify the current frame based on the recent temporal history. This enables continuous recognition of multiple actions in real world scenarios such as monitoring elderly patients at home [14]. However, there is an additional requirement in NUI applications to detect actions with low latency so the system can provide an appropriate response to the user with no apparent delay. For example, increasing the volume on a Smart TV by raising a hand should be detected with low latency to provide natural interaction.

Existing work has demonstrated that action points [11], temporal anchors within the course of the action are important for evaluating the latency of the detection. An action point is a single pose that can be clearly and easily identified as a representative of an action. Several, sliding window approaches for online action recognition have been validated using action points [9], [15], [16]. Fothergill et al. [9] used fixed size sliding windows on the streaming data and performed the classification by a Random Forest. Similarily, Bloom et al. [15] used a fixed size sliding window and perform the classification by AdaBoost. However, the fixed size of the sliding window in both approaches is a source of classification error due to execution rate variations. To address this Zhao et al. [16] optimise the size of the segment during their feature extraction using a DTW variant for subsequence matching. However, as these methods were tested with temporally separated actions their ability to robustly detect compound actions is unclear. Especially as AdaBoost which achieved good performance on relatively simple actions [17] but when applied to more complex actions performance dramatically decreased [4].

Manual labelling of action points is possible in complex datasets as they represent the most significant part of the action, however subsequently automatically selecting a sequence of training examples around the point leads to inconsistencies. Firstly, as some actions have long duration such as defending (see Figure 2), later samples of the current action will be incorrectly selected as negative samples. Secondly, samples from another action class may be incorrectly selected due to the close proximity of neighbouring actions (see Figure 2). The first problem has been overcome by action segments [4] which incorporate the duration of the most significant part of the action. The second problem has not yet been adequately addressed but could be alleviated by reducing the need for labelling.

Transfer learning [18] has been beneficial to many machine learning research areas, including classification, regression and clustering problems to reduce the need to collect and label training data. However, transfer learning applied to action recognition is a relatively new topic with limited research in the computer vision community. Transfer learning has been used for cross-view action recognition [19], [20] to recognise human actions from different views. In both cases the methods were tested offline on a multi-view dataset (IXMAS) [21], which comprised of simple actions with simple backgrounds so it has limited applicability to real world scenarios.

More significantly transfer learning has been used cross-dataset [8], [36] to harness lab datasets to facilitate real-world action recognition. The aim is to generalise action models built from a source dataset to a target dataset, to alleviate the problem of labelling complex sequences. The source dataset typically has a clean background and each video clip may involve only one type of action and a single person, which describes most lab collected datasets. In contrast, in the target dataset the background may be cluttered and there may be multiple people and multiple actions which may overlap temporally. Cross-dataset learning aims to adapt the existing classifier from a source dataset to a new target dataset, while requiring only a small or even no labelled samples in the target dataset. Ma et al. [8] built a model within a multi-task framework so the actions of one domain are associated with its own features. The general Schatten *p*-norm was applied to mine the shared components between the lab data and the real world data. The main advantage of their approach is the ability to share knowledge between the

two datasets even if they have different action categories. However, the method was tested offline with sequences containing just a single action. Cao et al. [22] combine model adaption and action detection into a Maximum a Posterior (MAP) estimation framework for action detection. The advantage of this approach over the previous method is that it can perform spatial-temporal detection of the action within a sequence. However, as a search for the optimal 3D sub-volume is performed across all frames in the target sequence this approach is also offline.

The approaches described so far are limited to single actions or multiple actions that are temporally separated. However, in NUI applications the user may wish to perform multiple actions in quick. This temporal merging of different actions results in complex poses comprising of movements from multiple actions. Hierarchical models have been successfully applied to pose estimation [23]–[27] to recover novel poses not present in the training dataset. Hierarchical models have also been applied to improve action recognition performance [28]. Following the popular bag-of-words approach several efforts constructed a hierarchical representation of local feature descriptors but as the temporal order is ignored they are not suited to many real-world problems. To overcome this Song et al. [29] propose hierarchical sequence summarisation to capture discriminative information at various temporal resolutions. However, as the testing was performed at the sequence level this approach is limited to offline action recognition.

2.1 Contributions

We propose a novel hierarchical transfer learning algorithm for online detection of compound actions for robust action recognition in natural user interface (NUI) applications. Specifically, transfer learning is employed to allow the tasks of action segmentation and modelling to be performed on a related but simpler dataset, combined with model adaptation to improve performance on a complex NUI dataset. We represent actions using a hierarchical human body model to allow independence between low-level body parts. Our novelty is to automatically weight each low-level body part based on their discriminative ability to detect specific actions. We propose hierarchical peak poses for low latency detection which provide the flexibility to recognise poses that are not in the source dataset. Hierarchical template matching is performed with Dynamic

Time Warping (DTW) to ensure execution rate invariance and we use a sliding window approach for online recognition. Evaluation on a public dataset with complex, realistic actions demonstrates that our approach outperforms existing methods in terms of accuracy and latency.

3. Methodology

The proposed method for online action recognition consists of two phases: an offline training phase and an online testing phase as illustrated in Figure 3.

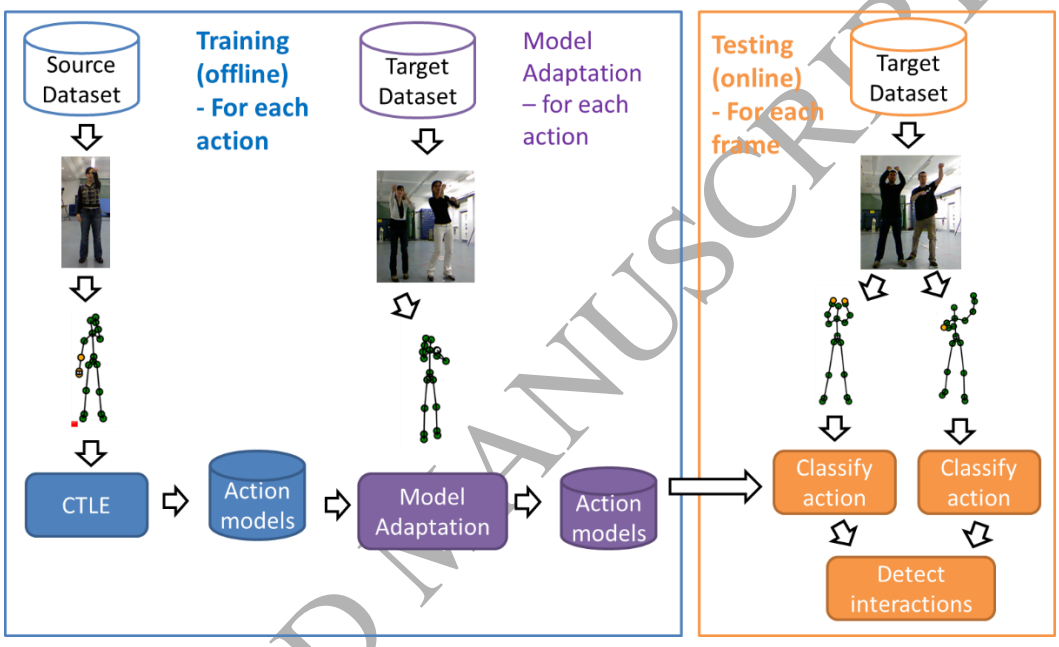


Figure 3 Methodology overview

We propose a novel hierarchical transfer learning algorithm for online detection of compound actions for fast and robust action recognition in natural user interface (NUI) applications. Our method is based on skeleton data, specifically joint angles which are viewpoint and anthropometric invariant and can be generated in real-time with a pose estimation method [30]. A key contribution is our hierarchical body model that can be automatically configured to detect actions based on the low level body parts that are the most discriminative for a particular action. Another key contribution is a transfer learning strategy to allow the tasks of action segmentation and whole body modelling to be performed on a related but simpler source dataset, combined with automatic hierarchical body model adaption on a more complex target dataset (as shown in Figure 3).

3.1 Training (source dataset)

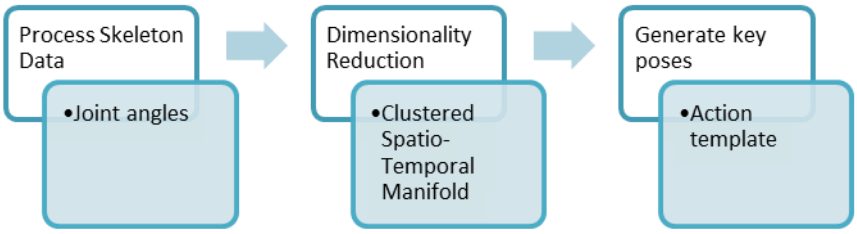


Figure 4 Training overview which is performed on the source dataset for each action

The training phase is based on our existing approach for online action detection [17] that achieved high accuracy and low latency for multiple actions that were separated temporally (see Figure 4). Our contribution is to adapt these action templates to detect compound actions by representing and detecting actions hierarchically. The two key stages in training, as published in our previous work [17] are dimensionality reduction and key pose generation. Dimensionality reduction of the skeleton data produces spatio-temporal manifolds which removes individual style whilst maintaining the temporal ordering of the poses. Clustering the manifolds and projecting the cluster centres back to the high dimensional space creates key poses. An individual key pose represents a generic pose from an action at a specific point in time and the sequence of these key poses represent the entire action (as illustrated in Figure 5). A major benefit of the clustering is that the number of key poses is significantly less than the original number of training poses which dramatically reduces the computation time and enables our approach to scale efficiently to much larger datasets.

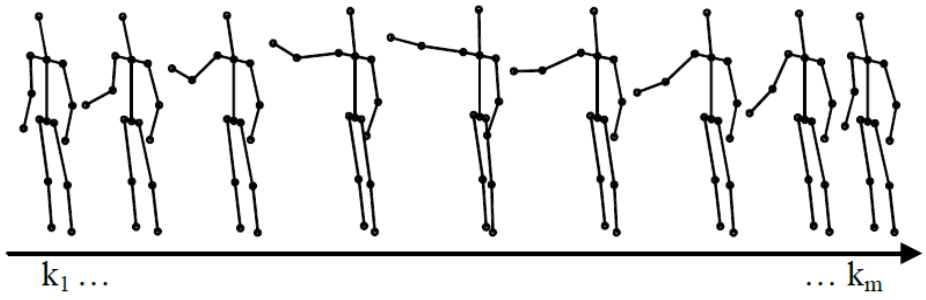


Figure 5 Right punch action template, consisting of key poses k_1 to k_m where m is the number of clusters [17]

The two stages are explained in detail below:

3.1.1. Dimensionality reduction

Stylistic variations are removed by learning a clustered spatio-temporal manifold (CSTM) for each action [17]. Given a set of training poses from the source dataset $X = \{x_i\}_{(i=1\ldots n)}$, $x_i \in \mathbb{R}^D$, distributed in a high dimensional space, Temporal Laplacian Eignemaps (TLE) [31] discovers their low dimensional representation $X' = \{x'_i\}_{(i=1\ldots n)}$, $x'_i \in \mathbb{R}^d$ where $d \ll D$ by combining two neighbourhood graphs. Temporal neighbours are the closest points in the sequential order and spatial neighbours are the geometrically similar neighbours. These neighbour relations are used in the construction of two graphs where any two vertices are connected when a neighbour relationship exists between these points. Neighbourhood connections defined in the Laplacian graphs place neighbours from the high dimensional space nearby in the embedded space. Consequently, the temporal neighbours preserve the temporal structure and the spatial neighbours reduce style variability by aligning the time series in the embedded space (see Figure 6).

Clustering is then performed on the embedded space to reduce computation time by removing redundant poses. *k*-means [32] is applied to cluster the n low dimensional points X' into m clusters $C = (c_j)_{(j=1\ldots m)}$, $c_k \in \mathbb{R}^d$, where $m \ll n$.

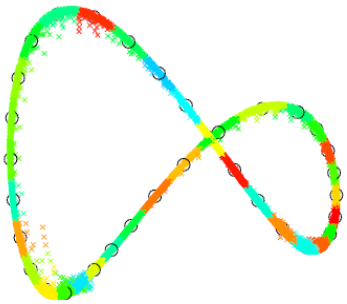


Figure 6 Clustered Spatio-Temporal manifold with the low dimensional points plotted (x_i), coloured based on the cluster to which they belong and the cluster centers (c_j) as black circles [17].

3.1.2. Key pose generation

Key poses remove redundant information to improve classification accuracy and reduce the computational latency of action detection [14], [17]. To generate key poses we follow the method proposed in [31] that uses the training set $M = \{x_i, x'_i\}_{(i=1\ldots n)}$ to learn a Radial Basis Function Network (RBFN) that

represents the mapping between the embedded and the high dimensional space [31]. Then using the RBFN mappings the cluster centers are projected into the high dimensional space to generate new poses that are a direct representation of the average poses. The implicit temporal order in the low dimensional space can be extracted from the training data to order the corresponding key poses $K = \{k_j\}_{(j=1\ldots m)}$ to create action templates (K) for each action as illustrated in Figure 5. Action templates are the high dimensional representations of the clustered spatio-temporal models and inherit their advantages, including style invariance and compactness.

3.2. Model Adaptation (target dataset)

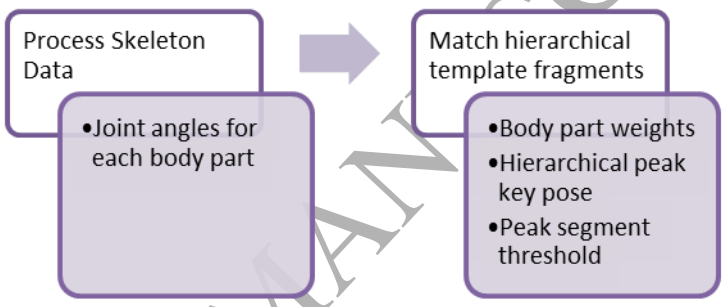


Figure 7 Model Adaptation overview which is performed on the target dataset for each action

To detect compound actions such as those performed in NUI applications we propose a hierarchical template matching algorithm (see Figure 7). Representing actions using a hierarchical model of human body allows independence between the low-level body parts $B = (b_l)_{(l=1\ldots L)}$ (as illustrated in Figure 8). Each low-level body part is represented by joint angles. Our contribution is to automatically weigh each low-level body part based on their discriminative ability to detect specific actions. Weighting the individual low-level body parts, creates flexible body part configurations at different levels of a normal body hierarchy e.g. whole body, upper body or right arm and atypical combinations such as right arm and left leg.

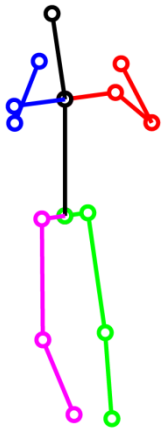


Figure 8 Low level body parts: the skeleton is divided into low level body parts, right arm (red), left arm (blue), right leg (green), left leg (pink) and torso (black).

The action peak is a fundamental concept of the proposed approach which we define as the segment in time when the goal of the action is being satisfied. For example, in a boxing game the aim of the punch is to hit the opponent which is being fulfilled when the arm is maximally extended as shown in Figure 9. The peak poses in the training data of the target dataset are manually labelled with an action label, there must be at least one frame labelled as the peak pose for each action instance. If the action peak has duration, as in the case of the defense action there will be multiple sequential labelled frames.

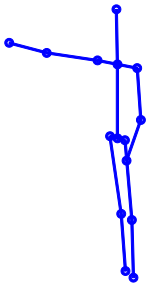


Figure 9 Action peak for right punch action

There are three main steps to adapt the action templates learnt from the source dataset for hierarchical template matching: learning the most discriminative body part combinations, detecting the most representative hierarchical peak key pose and optimising the peak segment threshold.

All three steps use exemplar matching between the peak poses in the target dataset training poses and the action templates to find the optimum matching parameters. To incorporate the temporal history of the action and increase the robustness of the matching process sequences of poses are matched rather than

single poses. To extract a fragment F from a sequence of poses $S = (s_1, s_2, \dots s_G)$ Eq. 1 is used:

$$F(S, i) = (s_{i-s}, s_{i-s+1}, \dots s_i) \tag{1}$$

where, i is the pose index, s is the number of poses in the fragment and G is the number of poses in sequence S and the conditions $i > s$ and $i \leq G$ are satisfied.

DTW [33] is a well-known algorithm for determining the similarity of time-series data that allows "elastic" transformation to gain execution rate invariance. The similarity of two series of poses, the query sequence $Q = (q_1, q_2, \dots q_U)$ and the reference sequence $R = (r_1, r_2, \dots r_V)$ can be computed using the standard DTW distance metric using Eq. 2.

$$DTW(Q, R) = \min\{c_p(Q, R), p \in P^{U \times V}\} \tag{2}$$

Where c_p is the global cost function associated with a warping path $p = (p_1, \dots, p_H)$ and c is the local cost function, which is the Euclidean distance between two poses, which will be small if the poses are similar to each other:

$$c_p(Q, R) = \sum_{h=1}^{H} c(q_{uh}, r_{vh}) \tag{3}$$

In our previous approach [17] the DTW distance was computed for the whole body. To increase flexibility we propose a hierarchical DTW distance measurement ($HDTW$):

$$HDTW(Q, R, W) = \sum_{l=1}^{L} DTW(Q_l, R_l) W_l \tag{4}$$

For two series of poses, the query sequence Q and the reference sequence R, the similarity of low level body parts l is computed independently using the standard DTW distance metric. A weighted combination $W = (w_l)_{(l=1\dots L)}, w_l \in (0,1)$ of the low level body part distances provides a discriminative distance metric for compound actions.

3.2.1. Body Part Combinations

The most discriminative body part combinations for each action are discovered by maximising the ratio of intra-class matches between the labelled peak poses in the target dataset training data and the action templates. This procedure is repeated

for all body part combinations, so for computational efficiency we selected binary weights, $w_l \in (0,1)$ for each of the low level body parts which results in 2^L permutations. For each permutation ε, the intra-class ratio ρ is computed by the number of intra-class matches μ over the number of total training instances in the target dataset n^y. The intra-class matches are counted for each action by exemplar matching between the peak poses from the target dataset training data and the key poses from all the action templates. For each action a, if the closest matching action template is the same action this is counted as an intra-class match. The maximum intra-class ratio represents the most discriminative body part combination for each action, as illustrated in Figure 10 and summarised in Algorithm 1.

Algorithm 1 Learn the most discriminative weights for each action

Input: Given a set of training poses from the target dataset $Y = \{y_i\}_{(i=1\ldots|Y|)}$, with manually selected peak poses from Y represented by their indices $I^a = \{i_p^a\}_{(p=1\ldots|I^a|)}$, where $i_p \in 1 \ldots |Y|$ and the superscript denotes a set of action templates $K^a = \{k_j\}_{(j=1\ldots m)}$, where m is the number of clusters.

1. For each action, $a = 1{:}\,A$
   1.1. For each permutation, $\varepsilon = 1{:}\,2^L$
      1.1.1. Initialise $\mu = 0$
      1.1.2. For each peak pose, $p = 1{:}\,|I^a|$
         1.1.2.1. Extract the peak pose fragment, $F^Y = (Y, i_p^a)$ using Eq. 1
         1.1.2.2. $a^* = \min_{a' \in A} HDTW(F^Y, K^{a'}, W_\varepsilon)$ using Eq. 4
         1.1.2.3. If $a^* = a$
            1.1.2.3.1. Intra-class match so increment μ
      1.1.3. Compute intra-class ratio, $\rho_\varepsilon^a = \frac{\mu}{|I^a|}$
   1.2. Select the most discriminative weights, $W^a = \arg\max_\varepsilon \rho_\varepsilon^a$
   1.3. Output the weights for this action, W^a

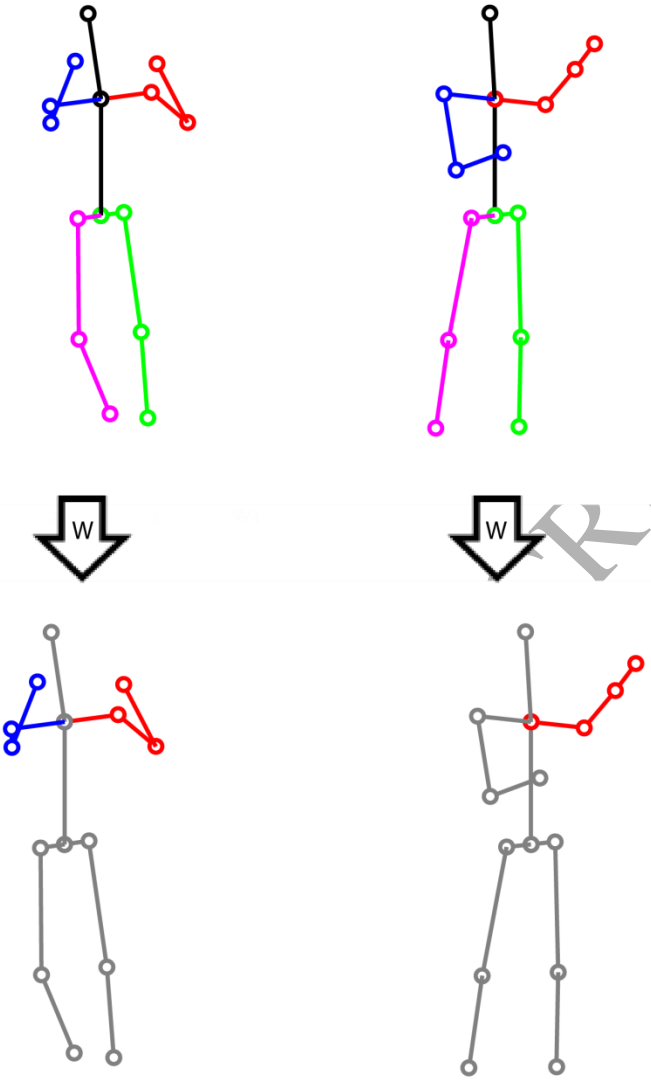


Figure 10 Body Part Combinations: The weights (W) are optimised for each action based on their ability to discriminate complex actions in the target dataset. The bottom skeletons show potential body parts configurations for the defend (left) and right punch (right) actions.

3.2.2. Hierarchical Peak Key Pose

In our previous work on simple actions, peak key poses were proposed as the generic representation of peak poses in the training data and were automatically selected from the key poses by exemplar matching with the whole body [17]. To increase robustness on compound actions we propose hierarchical peak key poses. Hierarchical peak key poses are also automatically selected from the key poses but the exemplar matching is performed using the most discriminative body parts rather than the whole body. The hierarchical peak key poses are selected as follows: for each action and for each peak pose in the target dataset training data,

the best matching key pose is found (as shown in Figure 11). A hierarchical peak key pose can be represented by its index j^a in the action template. The best matching index j^* is found by minimising the distance between the peak pose fragments F^{Y} and the key pose fragments F^{K} using the most discriminative body part combination for each action. The hierarchical peak key pose for the action is the key pose that has the maximum number of matches, as summarised in algorithm 2.

Figure 11 Hierarchical template matching: peak pose (left), best matched key pose (right)

Algorithm 2 Learn the hierarchical peak key pose

Input: Given a set of training poses from the target dataset $Y = \{y_n\}_{(n=1\ldots|Y|)}$ with manually selected peak poses from Y represented by their indices $I^a = \{i_p^a\}_{(p=1\ldots|I^a|)}$, where $i_p \in 1\ldots|Y|$ and the superscript denotes a set of action templates $K^a = \{k_j\}_{(j=1\ldots m)}$ with weights W^a:

1. For each action, $a = 1{:}\,A$
   - 1.1. Initialise $J = \{0\}_{(1\ldots m)}$
   - 1.2. For each peak pose, $p = 1{:}\,|I^a|$
     - 1.2.1. Extract the peak pose fragment $F^{\mathrm{Y}} = (Y, i_p^a)$ using Eq. 1
     - 1.2.2. Find the best matching hierarchical key pose index,

       $j^* = \arg\min_{j\in 1\ldots m} \sum_{l=1}^{L} HDTW\left(F_l^{\mathrm{Y}}, F_l^{\mathrm{K}}, W_l^a\right)$, where $F_l^{\mathrm{K}} = (K_l^a, j)$
     - 1.2.3. Increment J_{j^*}
   - 1.3. Output the hierarchical key pose index $j^a = \arg\max \; J$

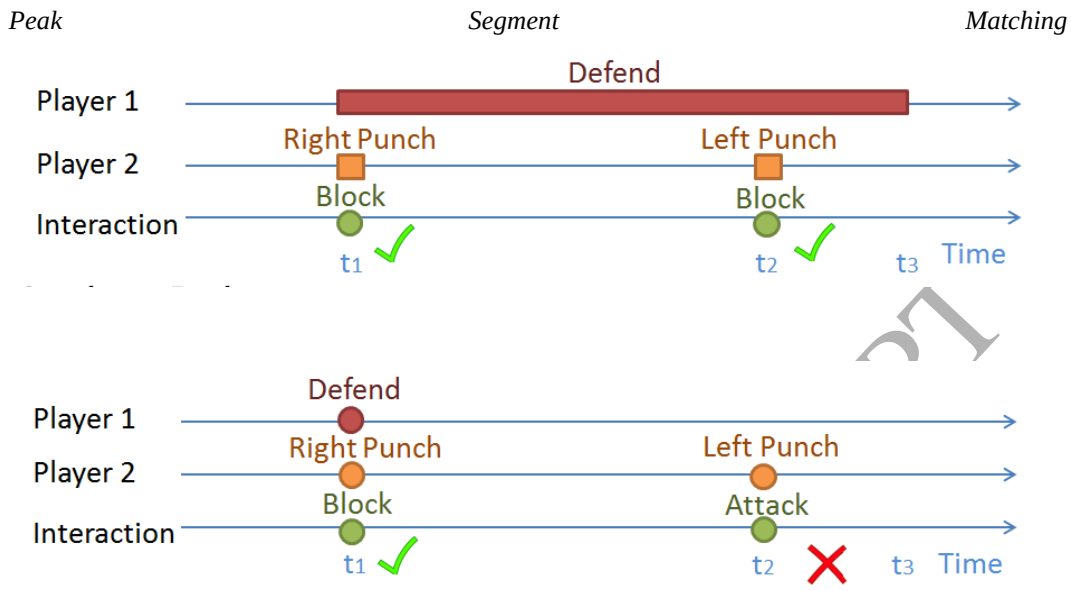


Figure 12 (Top) Interaction detection based on action segments which correctly detects actions with long duration. (Bottom) Interaction detection based on action points which only works if both actions occur at the same time and incorrectly detects interactions if an action has a long duration.

Some existing methods for online action recognition detect the action as a single point in time [9], [17] whereas others incorporate the duration of the action [14], [34]. The duration of the action is important for subsequently detecting interactions between multiple players in a sports game [4] and illustrated by Figure 12.

Peak key poses [14] were limited to detecting a single temporal point so we introduce a threshold τ to incorporate the duration of the peak. Similar to [14], [34] we introduce a threshold τ for action detection but instead of specifically learning a threshold for each action we learn a single threshold for all actions. Confining, the threshold to a single parameter reduces the time taken to adapt the model and this time will not increase even if more actions are considered, providing scalability to larger datasets.

The threshold τ and fragment size s are learnt on the training part of the target dataset by optimising the action point metric F1 [11] with our hierarchical template matching algorithm (summarised in Algorithm 3) but using the training data from the target dataset rather than the testing data.

3.3. Testing (target dataset)

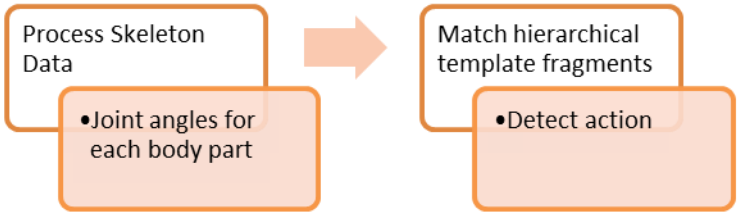


Figure 13 Testing overview which is performed on the target dataset

We propose a hierarchical template matching algorithm with a temporal sliding window for online action recognition (summarised in Algorithm 3). For each new frame the sliding window buffer is updated and compared with learnt exemplars. The minimum hierarchical DTW distance to the nearest neighbour is used to detect the action (see Figure 13).

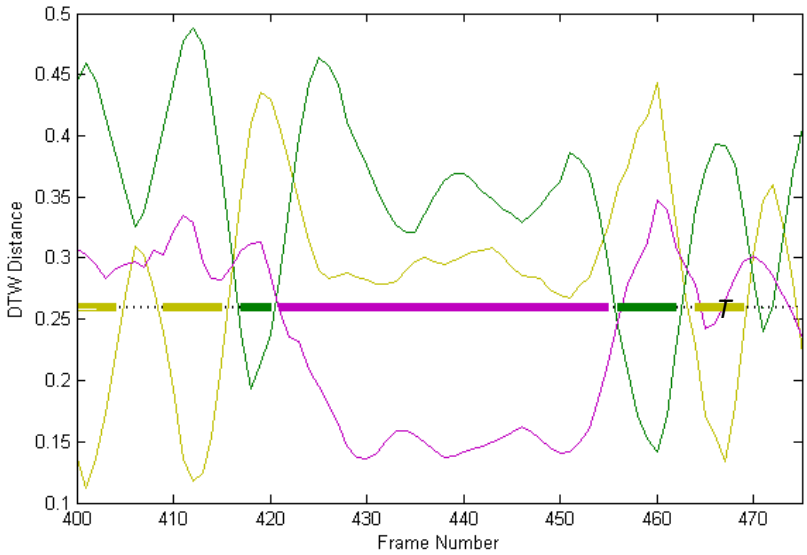


Figure 14 Normalised hierarchical DTW distances: the lowest value represents the most similar action, where this value is lower than the threshold τ it represents the detected action. The right punch is displayed in yellow, left punch displayed in green and the defend in magenta.

The hierarchical matching process is performed using DTW to ensure execution rate invariance. The normalised hierarchical DTW distances d^*, are recorded for each frame as illustrated in Figure 14. To detect actions in real-time we compare the lowest hierarchical DTW distance at each frame with a threshold τ. τ discriminates which pose fragments are most similar to the peak key pose

fragment. Therefore, whilst pose fragments are similar to the peak key pose fragment ($d^* \leq \tau$) the action is at its peak, as shown by the coloured segments on Figure 14. Before and after the peak, the pose fragments will be less similar ($d^* > \tau$) and therefore the action is not considered at its peak.

Algorithm 3 Online hierarchical template matching

Input: Given a set of testing poses from the target dataset $Z = \{z_i\}_{(i=1\ldots|Z|)}$, a set of action templates $K^a = \{k_j\}_{(j=1\ldots m)}$, with weights W^a, hierarchical peak key poses indices j^a, the fragment size s, and distance threshold τ:

1. For each testing pose, $i = 1{:}\,|Z|$
   1.1. Add the current test pose to the test fragment, $F^{\mathrm{Z}} = F^{\mathrm{Z}} \cup z_i$
   1.2. If $i \geq s$
      1.2.1. $F^{\mathrm{Z}} = F^{\mathrm{Z}} \setminus z_{i-s}$
   1.3. For each action, $a = 1{:}\,A$
      1.3.1. Extract the key pose fragment, $F^{\mathrm{K}} = (K^a, j^a)$ using Eq. 1
      1.3.2. Compute $HDTW(F^{\mathrm{Z}}, F^{\mathrm{K}}, W^a)$ using Eq. 4
   1.4. $d^* = \min_{a \in A} HDTW(F^{\mathrm{Z}}, F^{\mathrm{K}}, W_a)$
   1.5. If $d^* < \tau$
      1.5.1. $a^* = \arg\min_{a \in A} HDTW(F^{\mathrm{Z}}, F^{\mathrm{K}}, W^a)$
      1.5.2. Output "Action a^*"
   1.6. Else, output "No action"

One of the advantages of using clustering to identify peak poses is that the computational time is independent on the size of the training dataset, although it is linearly dependent on the number of actions. In case of many actions, a parallel implementation, i.e. one thread per action, would achieve real-time performance.

4. Experiments

In this section we present experiments to evaluate the ability of our online action recognition method to improve accuracy at low latency in complex scenarios.

4.1. Datasets

The performance of our algorithm is evaluated using publicly available datasets designed specifically for real time action recognition: G3D [10], MSRC-12 [9] and G3Di [4]. All datasets contain multiple actions in each sequence in a controlled indoor environment with a fixed camera, a typical setup for NUI applications. Both datasets provide sequences of skeleton data captured using the Kinect pose estimation pipeline at 30fps. However, G3D contains scripted actions which are temporally well separated whereas G3Di was captured using a gamesourcing approach where the users were recorded whilst playing computer games and consequently contains more complex actions which overlapping temporally. The G3Di also contains noisier skeleton data than G3D as there was interference from multiple Kinects during the recording, making it more realistic of a home scenario where there may be interference from the sunlight.

The G3D dataset contains 10 subjects performing 20 gaming actions grouped into seven categories. The fighting category was selected as it has the same actions as the G3Di boxing category although there are substantial variations in execution rate as well as personal style between these two datasets due to the different recording environments. The G3D fighting category contains five gaming actions: right punch, left punch, right kick, left kick and defend.

The MSRC-12 dataset comprises of 30 people performing 12 gestures. These gestures are categorized into two categories: iconic and metaphoric gestures. The iconic gestures directly correspond to real world actions and represent first person shooter (FPS) gaming actions. There are six FPS gaming actions: crouch, shoot, throw, night goggles, change weapon and kick. Whereas metaphoric actions represent abstract concepts for manipulating a music player e.g. raise volume of the music. The dataset was obtained using different instruction modalities and the modality that produced the most accurate results was video + text so we will use this particular subset of the dataset.

The G3Di dataset contains 12 people split into 6 pairs. Each pair interacted through a gaming interface showcasing six sports: boxing, volleyball, football, table tennis, sprint and hurdles. Boxing is a competitive sport and the interactions can be decomposed by an action and counter action. The boxing actions were right punch, left punch and defend and the interactions between the players are shown in Table 1. The total number of action and interaction instances used for our experiments is shown in Table 2.

Table 1 Gaming interactions for the boxing scenarios in G3Di.

Sport	*Action*	*Counter Action*	*Interaction*
Boxing	Right Punch	Defend	Block
	Left Punch	Defend	Block
	Right Punch	Other	Attack
	Left Punch	Other	Attack
	Right Punch	Right Punch	Attack
	Right Punch	Left Punch	Attack
	Left Punch	Left Punch	Attack

Table 2 The total number of action and interaction instances used from each dataset

Dataset	*Action Classes*	*Interaction Classes*	*Subjects*	*Action / Interaction Instances*	*Frames*
G3D (Boxing)	5	NA	10	150	12,870
MSRC-12 (Iconic Gestures)	6	NA	10	502	4782
G3Di (Fighting)	3	2	12	317 + 257 = 574	6784

4.2. Skeleton Data

Joint angles are viewpoint and anthropometric invariant and can be generated in real-time with a pose estimation method [30]. More specifically, the skeleton poses are first normalised and then the three angles defining each joint position are computed and represented by a 4-D quaternion. The skeleton is parameterised as a high dimensional feature vector by concatenating quaternions for all joints. For each pose 13 quaternions are calculated so each feature vector has 52-dimensions (see [14] for more details).

4.3. Comparative Study

The following is a brief introduction of the comparison algorithms in our experiments:

- AdaBoost: AdaBoost has shown high accuracy and low latency for online action recognition [5], [17]. AdaBoost was trained on the source dataset and the parameters: the number of training frames around each peak pose the sliding smoothing window size were optimised on the training part of the target dataset and the method was evaluated on to the target testing data.
- Clustered Spatio-Temporial Manifolds (CSTM): is a state-of the art approach for low latency online action recognition [17]. CSTM was trained on the source dataset and the parameters: the template size and the stream size and the peak pose detector were optimised on the training part of the target dataset and the method was evaluated on to the target testing data.
- Hierarchical Transfer Segments (HiTS): The proposed method in this paper, a version of CSTM extended for transfer learning, allowing knowledge to be transferred from simple actions in a source dataset to complex actions in a target dataset by adapting the learnt models with a hierarchical pose representation. The parameters: peak segment matching threshold (τ=0.22) and fragment size ($s = 7$) were optimised on the training part of the target dataset and the method was evaluated on to the target testing data.

For all the above experiments we performed leave one-person out cross validation on the target dataset; each cross validation fold was trained on 11 subjects and tested on the remaining subject.

4.4. Performance Metrics

Evaluating of action recognition algorithms has previously been done in isolation, focusing historically on high accuracy and more recently also on low latency. However, in reality most actions form part of an interaction where the duration of the action is important. To test our proposed algorithm in a realistic context we employ the interaction detection and evaluation framework [4] and the action point metric [11] which is the most commonly used metric for online action recognition.

4.4.1. Action Point Metric

For evaluation we use an existing latency-aware performance metric for based on temporal anchors known as action points [11]. For a specified amount of latency (Δms) the action point F1-score determines whether a detection made at time t_p for action a is correct in relation to a ground truth action point at time t_g by using the following formula:

$$\Phi(t_p, t_g, \Delta) = \begin{cases} 1 & \text{if } (\,|\, t_g - t_p \,|\, \leq \Delta) \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

For a specified amount of latency (Δms) the precision and recall are measured for each action and combined to calculate a single F-score.

$$\text{F1} - \text{score}(a, \Delta) = 2\frac{\text{prec}_a(\Delta)\ \text{rec}_a(\Delta)}{\text{prec}_a(\Delta) + \text{rec}_a(\Delta)} \quad (6)$$

As online action recognition algorithms need to detect multiple actions, the mean F-score over all actions is used, defined as:

$$\text{Average F1} - \text{score}(A, \Delta) = \frac{1}{|A|} \sum_{a \in A} \text{F1} - \text{score}(a, \Delta) \quad (7)$$

4.4.2. Interaction Detection Framework

The Interaction Detection Framework [4] enables online interaction recognition between multiple people by detecting their individual actions independently and

combining them by a set of interaction rules to infer the interaction. This modular approach is applicable for NUI and enables interaction between people that are not in the same physical location. Actions from different people are detected independently. At each frame, these detections are combined to infer the current interaction. The interaction rules include the valid combinations of actions (as depicted in Table 1) together with timing constraints. The action (*a*) and counter action (*ca*), are checked at each frame together with a timing constraint (*f*) to detect interactions in real time using Eq. 8. The timing constraint depends on the scenario, for example all the interactions in boxing are instant (f = 0), the action and counter action co-occur.

$$\psi\,(a_s, a_e, ca_s, ca_e) = \begin{cases} 1 \; if \; (a_s + f \leq ca_e) \; and \; (ca_s \leq a_e + f) \\ 0 \; otherwise \end{cases} \qquad (8)$$

Where s and e represent the start and end of the action segment respectively and $s \leq e$.

4.5. Results

Our method (HiTS) outperforms existing state of the art approaches for fast online action and interaction recognition, as shown in Figure 155. Both AdaBoost and CSTM show a significant drop in accurately detecting actions on the G3Di (Fighting) dataset in comparison with previously published results [17] on the G3D (Boxing) dataset. This is significant especially as the G3Di (Fighting) actions are a subset of the G3D (Boxing) actions but confirms our hypothesis that compound actions are more difficult to detect than multiple actions that are temporally well separated.

Additionally, we highlight the recognition accuracy for each category of action and interaction for a more detailed analysis of each method, as shown in Figure 16. A significant outcome is that even though CSTM [17] can detect all of the action categories it is unable to detect any interactions which are comprised of actions with duration, specifically the block interaction. In addition to showing the limitation of this approach it also highlights a weakness of the action point metric [11] which does not incorporate the duration of the action peak. Interaction detection is improved by our baseline method Peak Segment Matching (PSM) which instead of a binary decision for matching a peak key pose introduces a threshold which can detect the duration of the peak. The key contributions of this

paper are the hierarchical body model (HSM) and a transfer learning strategy (TSM). Individually, applied to our baseline method these contributions actually decrease the action and interaction recognition but together (HiTS) they form a powerful combination that significantly increases the action and interaction recognition, as shown in Figure 12. Intuitively, our hierarchical representation is only useful if adapted to the target dataset.

In this paper we are exclusively interested in action recognition approaches that are suitable for NUI applications. Research has shown that a delay of 100ms is not perceivable by the user [35]. Therefore, in this section we have only compared our method against online action recognition methods that are capable of fulfilling this requirement. Table 3 shows that all the methods we evaluated are capable of detecting actions with a low average latency of approx. 2 frames, which is equivalent to 66ms. We did not evaluate online action recognition methods with high latency (830-1500ms [16], 2000ms [14]) as they are better suited to other applications.

Table 3 A comparison of the average action latency

Method	Average Action Latency (frames)
AdaBoost	2.12
CSTM	2.00
PSM	1.60
TSM	1.41
HSM	1.94
HiTS	2.36

Figure 17 illustrates a typical failure case caused by noisy skeleton data at the action level resulting in an incorrect interaction to be inferred. The main limitation of our approach is that we only utilise the skeleton modality which is subject to interference from sunlight.

The dependency of the proposed transfer learning methodology on the amount of training data used from the target dataset is investigated. Specifically, Figure 18 demonstrates the action and interaction recognition performance (F1) for varying number of training subjects. The proposed method may achieve similar results to other competitor methods, i.e. around 0.6 and 0.4 F1 score for action and

interaction recognition respectively (see Figure 15) with almost half the training data from the target dataset, i.e. 6 subjects.

Regarding the template size *s* in theory it is possible to use different values in the matching process. However, in practice it was not computationally feasible to test all of these combinations so in our experiments we actually used a single parameter s which was learnt on the training part of the target dataset. Figure 19 shows how this parameter affects performance. This parameter does not model the duration of the action as the graph shows that even 3 frames (100ms) can accurately detect the action peak and overall performance is fairly consistent for higher values.

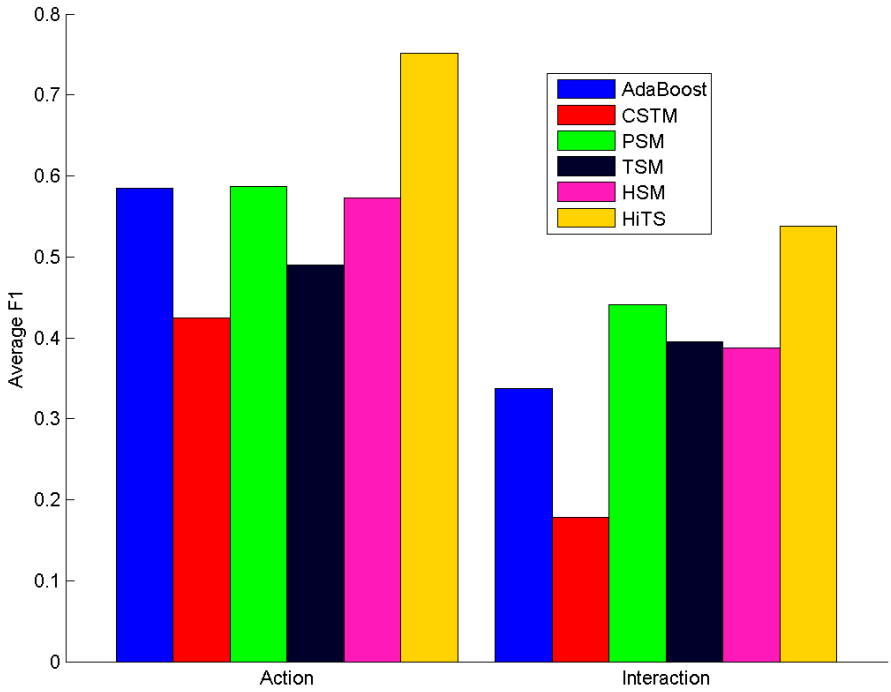


Figure 15 Performance comparison of the different approaches. Our method (HiTS) outperforms the others for both action and interaction detection.

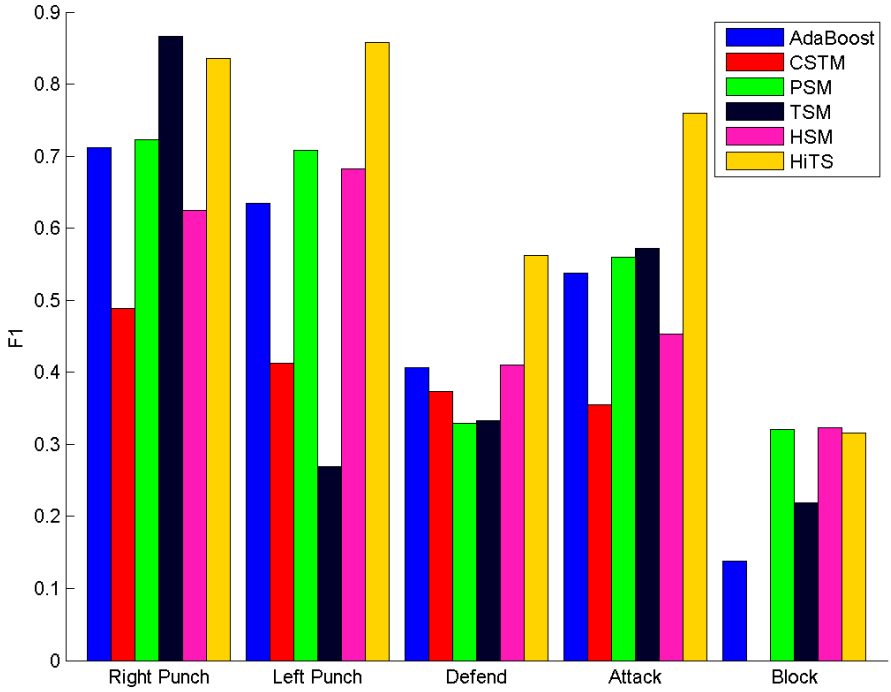


Figure 16 Action recognition results (left) and interaction recognition results (right) for each category of the G3Di (Fighting) dataset using different algorithms

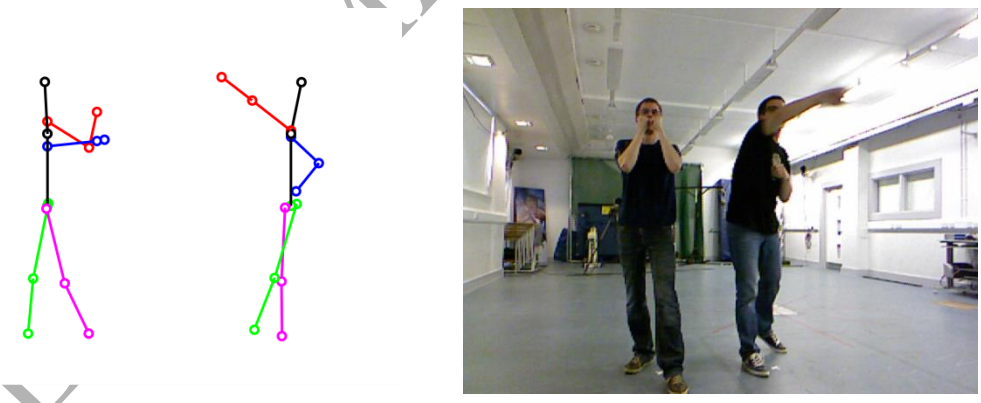


Figure 17 Example of a typical failure case caused by noisy skeleton data. The colour image (right) shows that this is a block interaction but our algorithm detects an attack interaction as the defend action is not correctly detected due to incorrect skeleton data for the player on the left. This instance will be penalised twice by the action point metric, firstly a FP for the attack and secondly a FN for the block.

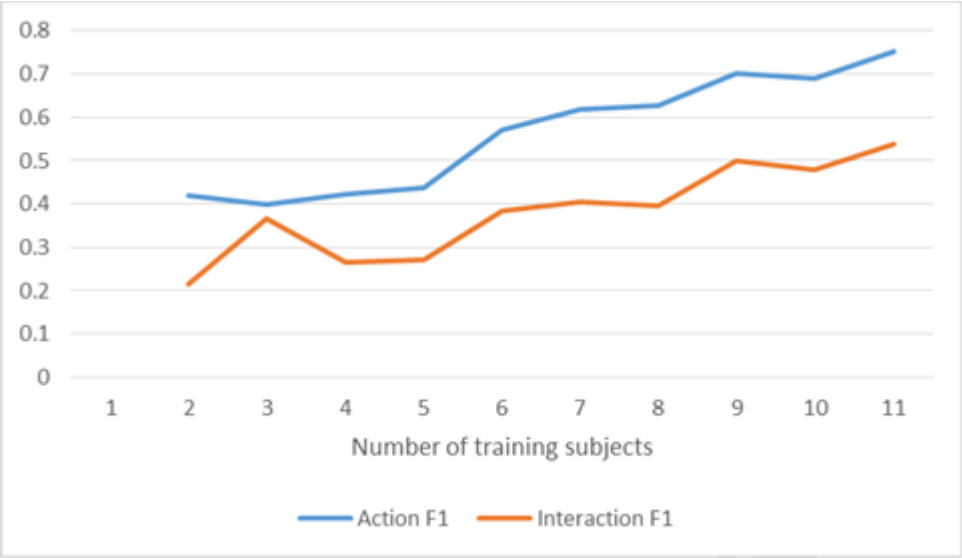


Figure 18 The relationship of the required number of training subjects and the obtained accuracy (F1 score) both for action and interaction analysis.

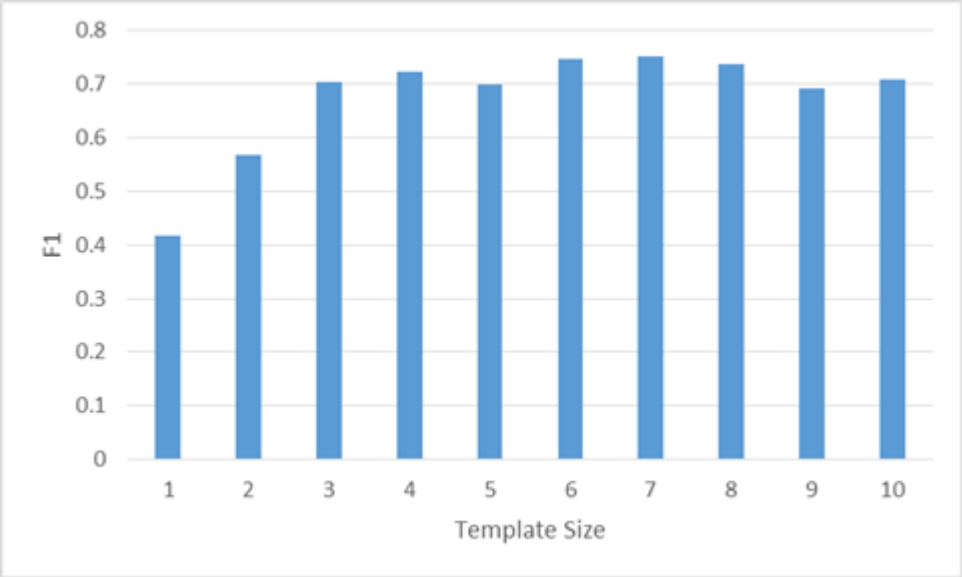


Figure 19 An example that indicates the relationship between the template size and the obtained accuracy (F1 score).

5. Conclusion

In this work we presented a novel hierarchical transfer learning algorithm for fast online action recognition. It overcomes the limitations of existing approaches by representing the human body hierarchically and learning the most discriminative body parts to detect compound actions. A transfer learning strategy was introduced to allow the tasks of action segmentation and whole body modelling to be performed on a related but simpler dataset. Combined with

hierarchical model adaptation on a more complex dataset to introduce independence between limbs and provide the flexibility to recognise poses that are not in the source dataset. Evaluation on a public target dataset that is more challenging and realistic than the source dataset shows our hierarchical transfer learning algorithm significantly increases performance at low latency. As the target dataset was recorded whilst users were actually playing a game the actions are more natural than subjects that are given instructions or restrictions and demonstrates the viability of our algorithm for use in real-world applications.

The limitation of our approach is that we only utilise the skeleton modality which is subject to interference from sunlight. Our future work is improve the robustness of our algorithm by fusing features from the depth or colour with our hierarchical skeleton features and evaluate its effectiveness using the G3Di multi-modal dataset.

References

[1] H. Liu, R. Feris, and M. Sun, "Benchmarking Datasets for Human Activity Recognition," in *Visual Analysis of Humans*, T. B. Moeslund, A. Hilton, V. Krüger, and L. Sigal, Eds. London: Springer London, 2011, pp. 411–427.

[2] A. Barbu, D. Barrett, and W. Chen, "Seeing is worse than believing: Reading people's minds better than computer-vision methods recognize actions," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2014, pp. 612–627.

[3] C. Schuldt, I. Laptev, and B. Caputo, "Recognizing human actions: a local SVM approach," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, 2004, vol. 3, pp. 32–36 Vol.3.

[4] V. Bloom, V. Argyriou, and D. Makris, "G3Di : A Gaming Interaction Dataset with a Real Time Detection and Evaluation Framework," in *European Conf. on Computer Vision Workshops (ECCVW)*, 2014.

[5] J. Liu, J. Luo, and M. Shah, "Recognizing realistic actions from videos 'in the wild,'" in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 1996–2003.

[6] I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld, "Learning realistic human actions from movies," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[7] M. D. Rodriguez, J. Ahmed, and M. Shah, "Action MACH A Spatio-temporal Maximum Average Correlation Height Filter for Action Recognition," in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'*, 2008, pp. 1–8.

[8] Z. Ma, Y. Yang, F. Nie, N. Sebe, S. Yan, and A. G. Hauptmann, "Harnessing Lab Knowledge for Real-World Action Recognition," *Int. J. Comput. Vis.*, vol. 109, no. 1–2, pp. 60–73, Apr. 2014.

[9] S. Fothergill, H. Mentis, P. Kohli, and S. Nowozin, "Instructing people for training gestural interactive systems," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2012, pp. 1737–1746.

[10] V. Bloom, D. Makris, and V. Argyriou, "G3D: A gaming action dataset and real time action recognition evaluation framework," in *Computer Vision and Pattern Recognition Workshop (CVPRW), 2012 IEEE Conference on*, 2012, pp. 7–12.

[11] S. Nowozin and J. Shotton, "Action Points: A Representation for Low-latency Online Human Action Recognition," *Techincal Rep.*, pp. 1–18, 2012.

[12] J. K. Aggarwal and M. S. Ryoo, "Human activity analysis: A review," *ACM Comput. Surv.*, vol. 43, no. 3, pp. 16:1–16:43, Apr. 2011.

[13] J. Han, L. Shao, D. Xu, and J. Shotton, "Enhanced computer vision with microsoft kinect sensor: a review.," *IEEE Trans. Cybern.*, vol. 43, no. 5, pp. 1318–34, Oct. 2013.

[14] A. Chaaraoui and F. Flórez-Revuelta, "Continuous Human Action Recognition in Ambient Assisted Living Scenarios," in *First International Workshop on Enhanced Living EnvironMENTs (ELEMENT)*, 2014, pp. 1–8.

[15] V. Bloom, V. Argyriou, and D. Makris, "Dynamic Feature Selection for Online Action Recognition," in *Human Behavior Understanding, Lecture Notes in Computer Science*, vol. LNCS, no. 8212, Switzerland: Springer International Publishing, 2013, pp. 64–76.

[16] X. Zhao, X. Li, C. Pang, X. Zhu, and Q. Z. Sheng, "Online Human Gesture Recognition from Motion Data Streams," in *ACM Multi-Media 2013*, 2013, pp. 23–32.

[17] V. Bloom, D. Makris, and V. Argyriou, "Clustered Spatio-temporal Manifolds for Online Action Recognition," in *Proceedings of the International Conference on Pattern Recognition (ICPR)*, 2014, pp. 3963–3968.

[18] S. Pan and Q. Yang, "A survey on transfer learning," *... Data Eng. IEEE Trans.*, vol. 22, no. 10, 2010.

[19] A. Farhadi and M. Tabrizi, "Learning to recognize activities from the wrong view point," *Comput. Vision–ECCV 2008*, pp. 154–166, 2008.

[20] J. Liu and M. Shah, "Cross-view action recognition via view knowledge transfer," *Comput. Vis. ...*, 2011.

[21] D. Weinland, R. Ronfard, and E. Boyer, "Free viewpoint action recognition using motion history volumes," *Comput. Vis. Image Underst.*, vol. 104, no. 2–3, pp. 249–257, Nov. 2006.

[22] L. Cao, Z. Liu, and T. Huang, "Cross-dataset action detection," *Comput. Vis. pattern ...*, 2010.

[23] Y. Wang, D. Tran, and Z. Liao, "Learning hierarchical poselets for human parsing," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2011, pp. 1705–1712.

[24] Y. Tian, C. Zitnick, and S. Narasimhan, "Exploring the spatial hierarchy of mixture models for human pose estimation," *Comput. Vision–ECCV 2012*, 2012.

[25] L. Raskin, M. Rudzsky, and E. Rivlin, "Using hierarchical models for 3D human body-part tracking," *Image Anal.*, pp. 11–20, 2009.

[26] J. Darby, B. Li, N. Costen, D. Fleet, and N. Lawrence, "Backing Off: Hierarchical Decomposition of Activity for 3D Novel Pose Recovery.," *BMVC*, 2009.

[27] A. Moutzouris, J. Martinez-del-Rincon, J.-C. Nebel, and D. Makris, "Efficient tracking of human poses using a manifold hierarchy," *Comput. Vis. Image Underst.*, Oct. 2014.

[28] L. Han, X. Wu, W. Liang, G. Hou, and Y. Jia, "Discriminative human action recognition in the learned hierarchical manifold space," *Image Vis. Comput.*, vol. 28, no. 5, pp. 836–849, May 2010.

[29] Y. Song, L. Morency, and R. Davis, "Action recognition by hierarchical sequence summarization," *IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 3562 – 3569, 2013.

[30] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, *Real-time human pose recognition in parts from single depth images*, vol. 2, no. 3. IEEE, 2011, pp. 1297–1304.

[31] M. Lewandowski, D. Makris, and J. Nebel, "Temporal Extension of Laplacian Eigenmaps for Unsupervised Dimensionality Reduction of Time Series," in *International Conference on Pattern Recognition*, 2010, pp. 161 – 164.

[32] T. Kanungo, D. M. Mount, N. S. Netanyahu, A. Y. Wu, and C. D. Piatko, "A Local Search Approximation Algorithm for k-Means Clustering," *Spec. Issue 18th Annu. Symp. Comput. Geom. - SoCG2002*, vol. 28, no. 2–3, pp. 89–112, 2003.

[33] P. Senin, "Dynamic Time Warping Algorithm Review," USA, 2008.

[34] I. Kviatkovsky, E. Rivlin, and I. Shimshoni, "Online action recognition using covariance of shape and motion," *Comput. Vis. Image Underst.*, vol. 129, pp. 15–26, Dec. 2014.

[35] S. K. Card, G. G. Robertson, and J. D. Mackinlay, "The information visualizer: An information workspace," in *Proc. ACM CHI*, 1991, pp. 181–188.

[36] Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, and Andrew Blake, Real-Time Human Pose Recognition in Parts from a Single Depth Image, IEEE CVPR 2011